DRESDEN UNIVERSITY OF TECHNOLOGY
FACULTY OF MATHEMATICS AND NATURAL SCIENCES

DEVELOPMENT OF TRAPPING STYLE CASSETTES

FOR

NEW GENE TARGETING STRATEGIES

PhD THESIS

A thesis submitted to Dresden University of Technology for the degree Dr.rer.nat

SENEM SIMSEK

24.2.1974, Ankara, TURKEY

Reviewers: Prof. Dr. A. Francis Stewart, Dresden
Prof. Dr. Harald von Melchner, Frankfurt
Dr. Frank Bucholz, Dresden

Date of submission : 02.07.2007

Date of defense : 15.10.2007

ABSTRACT

Because of shared physiological, anatomical and metabolical features with humans, mice have served for a long time as mammalian disease models. In particular, these last ten years have been the golden age for this favoured model animal. Human and mouse genome projects show that there is 95% genome homology. Spurred by this fact, research attention has shifted from reading these sequences to deciphering the functions of these genes. The 1980s saw the remarkable achievement of homologous recombination in mammalian cell culture systems. Later in the 1990s, innovative gene trapping strategies were developed to enabled random mutagenesis. Today, the goal is to generate more versatile tools to avoid limitations posed by these earlier mutagenesis strategies. Many public and private research centers have united with the aim of mutating all mouse genes. In order to achieve this mutagenesis, the first requirement is a set of practical and efficient viral or plasmid based vectors that can be used globally in the genome. This will be aided by advances in understanding of biological events such as gene transcription, recombination, and embryonic stem cell cycle. In addition, technical improvements such as vector development, precise cell culture assay, and recombinant DNA delivery will also be important. The vector design work in this PhD thesis encompasses 0.00001 % ofthese efforts but may to out to be highly relevant.

The main aim of this PhD project was to construct a versatile FLEx gene trapping cassette that efficiently generates precise multipurpose alleles. Multipurpose allele design allows genome many types of genomic modifications such as knock-outs, hypomorphic alleles, conditional alleles, protein tags and GFP (green fluorescent protein) fusions. Multipurpose alleles reduce both the work and cost associated with mutageneis. FLEx is a strategy for inversional site-specific recombination. During the early phase of the cassette development study, the core region was formed by basic pre-mRNA processing elements: a synthetic splice acceptor **(84)**; a synthetic polyadenylation site **(55)**; transcription termination signals, CoTC **(26,116)** and/or MAZ4 **(3,36).** This core was flanked by inversely oriented heterotypic lox site pairs, loxP **(53)** and lox5171 **(53)**. For conditional mutagenesis, the FLEx cassette first needed to be neutrally orientated. Following Cre induction, it switches to the mutagenic orientation and acts at the level of

RNA processing to terminate transcription downstream of the cassette. This application was novel in its use of mammalian RNA polymerase terminators, CoTC and MAZ4.

Refinement of the vector design was required to find optimal combinations of the transcription elements. In order to increase efficient pre-mRNA processing, polyadenylation and splicing events, we explored factors including the relative distance between elements in the cassette and the role of the DNA between splicing acceptor and polyadenylation signals. Several FLEx cassette series containing these elements in a various configurations were engineered and tested in the first intronic region of pCAGGs-EGFP-SV40-Neo. This is a plasmid vector consisting of a strong promoter, a hybrid of the CMV early immediate enhancer and the chicken β-actin promoter, followed by a β-actin intron driving the enhancer green fluorescent marker fused with the neomycin selection marker under the control of te SV40 promoter. This laborious design task gave rise to a suitable FLEx gene trapping cassette. We next used our vector design to generate a multipurpose allele of mouse Jarid1c, a gene on the X chromosome. For this we produced FLEx-blasticidin that included the CoTC transcription termination signal and a targeting construct for Jarid1c.

Consideration of RNA processing events and transcription termination allowed us to develop a practical gene trapping cassette that can be mutagenic in a global manner. Strategic placement of mutant and wild type lox sites into the cassette resulted in high Cre recombination capacity. The work here thus outlines an elegant strategy to combine two allelic variations, a conditional wild type like and knock-out, in one multi-purpose allele. Therefore, the FLEx gene trapping cassette will save research time in investigations of processes such as cancer by efficiently creating specific mutations.

1. INTRODUCTION

1.1 Current strategies for structural and functional characterization of genes during mouse development

1.1.1 Background on genome analysis: The goal of genome projects

The genome of an organism is its total genetic information written in DNA code, a universal language. In 1990, the Human Genome Project was conducted as a large multidisciplinary project with the ultimate goal of sequencing all human DNA. The stunning completion of this project opened Pandora's box, exposing the raw material that encodes a human. Importantly, this work was an initial step towards understanding complex human developmental systems and exploring factors engaged in serious diseases and disabilities.

Although a function has not been described for the majority of genes, we have a lot of numerical information that characterizes the human genome structure. It has been reported that the human genome is composed of roughly 3 billion nucleotides **(International Human Genome Sequencing Consortium 2004)**. Less than 2% of this covers protein-coding regions. According to the International Human Genome Consortium (2004), the total number of genes is estimated to be 20.000-25.000. Although they have been sequenced and precisely mapped, over 50% are still waiting to be mined for functional and expression data. Beside protein coding sequences and regulatory elements, the human genome is also composed of regions, previously termed junk DNA, that do not code for proteins. These regions make up the vast majority, approximately 97%, of the human genome **(Diagram 1).** Even tough no direct functions have been assigned to these regions, they may affect chromosome structure and dynamics. It is crucial to understand the contribution of non-coding DNA to biological processes such as evolution.

Determining the functions of individual genes, regulatory sequences and junk DNA, as well as profiling their organization and interactions *in vivo*, is now the primary task of biologists. Knowledge of an organism's genome sequence is therefore a catalyst for other approaches to interpret genome function. Such interpretation requires time, expertise, and interdisciplinary collaboration. The new field of comparative and functional genomics will be the main force in revealing genome function. Genome sequencing is not limited to humans. Model organism genome projects promote comparative genome analysis as a powerful method to identify genes and other novel regulatory factors involved biological processes such as chromosome integration, and transcriptional regulation **(49,63,65,105)**. Functional genomics is the study of function related aspects of genome, gene transcription, translation factor analysis involved in these processes **(113)**. Together, comparative and functional genomics bring deeper insight to the human genetic script via gathering information about homologous gene and regulatory elements. In the near future, this information will fill gaps in our understanding of the molecular mechanisms and pathways responsible for normal mammalian development. It will also accelerate the development of strategies for diagnosis, treatment, and prevention of genetic diseases and disorders **(72)**.

1.1.2 Background on mouse as a model organism

Genetic mechanisms and cellular pathways are similar across species, including humans. Model organisms are often used in experimental systems designed to elucidate the fundamental mechanisms of developmental processes that underlying normal and pathological conditions. Depending on the specific area of research, yeast, worm, fly and/or mouse can be used as a model system to study human genes and to explain critical biological functions such as the cell cycle. *S.cerevisia, A.thaliana, D.melanogaster, C.elegans* and *Mus musculus* (house mouse) are the five organisms selected for the Human Genome project **(72)**. One of them, *Mus musculus*, has been found to be most promising for future studies. In addition to being small and easy to house, short generation time makes the mouse a good model animal. In April 2002 the Whitehead Institute released the complete mouse genome data. This sequence was then selected by

the Mouse Genome Sequencing Consortium to serve as the reference genome. This assembly contained approximately 22.444 mouse genes, 75% of them having a human genome counterpart. In summary, the high degree of similarity to human gene sequences, and human-like physiological and reproductive features, make the mouse an excellent model system. One other great advantage of using mice is that genomic manipulations can be used make more sophisticated mouse models. Finally, certain mouse models recapitulate human diseases such as lissencephaly, a neurological disorder **(52)**. Thus the mouse can serve as a model system for studies of immunology, neurobiology, reproduction, cancer, and epigenetics.

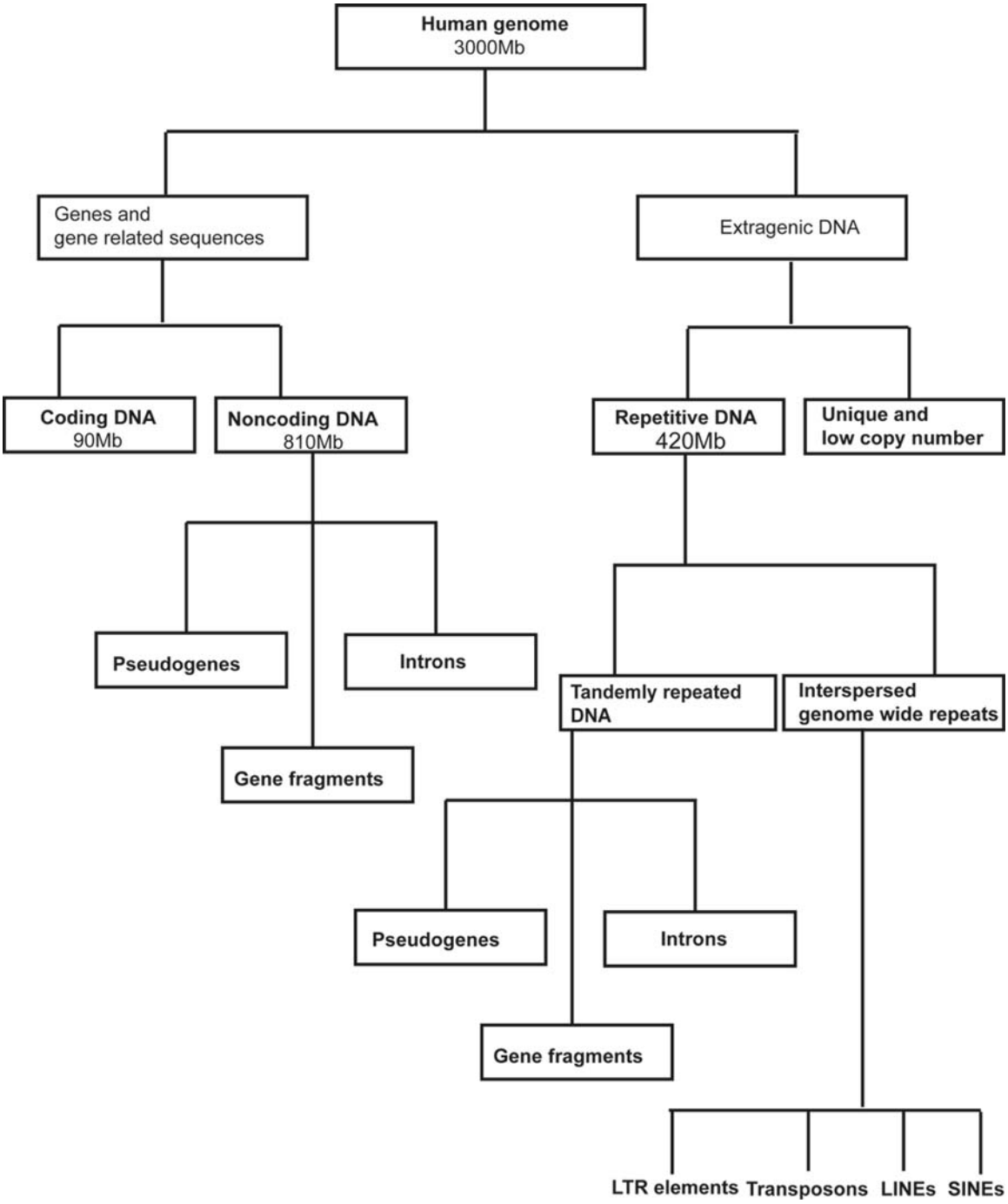


Diagram1 Numerical illustrations of human genome structure

(Strachan and Read 1996)

1.1.3 Basic tools for functional analysis of the mouse genome

In some cases, we can not determine gene functionality in mammalian systems by studying the DNA sequence or comparing sequences from two organisms. Therefore, the current challenge in genome projects is to develop reliable and efficient scientific approaches and methodologies to investigate gene function in mammalian development.

Mutagenesis is an efficient experimental approach to determine mammalian gene function by revealing the physiological consequences of specific gene mutation on the model organism. At the beginning, spontaneous mutagenesis based approaches were used for explore gene function in the mouse model. However, this was limited by low mutagenesis frequency ($5x10^{-6}$/locus). Although different mutation types, from point mutations to large chromosomal rearrangements, were produced and caused visible phenotypes, gene function could not be investigated in a well-organized, defined manner. Following this approach, X ray and chlorambucil approaches were invented to make mutants in a more regulated way These two approaches cause a wide range of chromosomal rearrangements, however, they affect more than one gene and create ambiguity in identification of target gene and its function. This led to the evolution of other approaches to design ideal mouse models for elucidating gene function. We can classify these classical mutagenesis technologies as: (1) phenotype driven mutagenesis, largely achieved by ENU chemical mutagenesis, and (2) genotype driven mutagenesis by gene trapping and gene targeting strategies. In addition to these, RNAi technology that inhibits gene transcription is promising to illuminate mammalian gene function.

Mutagenesis approaches take different routes to unravel genome functionality and each has advantages and disadvantages **(Table 1)**.

Summary of Basic Mutagenesis Approaches			
	Definition	***Advantages***	***Disadvantages***
Phenotype driven chemical mutagenesis	Generation of point mutations by alkylating agent ENU.It creates mostly partial loss of function and gain of function allele in germ cells	Generate allelic series of a single gene No prior assumption for genes and pathways	Random mutagenesis-unpredictable phenotypes * Difficult to identification of DNA regions responsible for desired phenotype
Gene trapping-genotype drive mutagenesis	Mutagenesis by random insertion of promoterless reporter gene flanked by splicing acceptor and polyA signal	Highthroughput for loss of function studies Rapid identification of distrupted gene and less time cost	Random fashion mutations-unpredictable phenotypes
Gene targeting-genotype driven mutagenesis	Targeted mutagenesis by homologous recombination between endogeneous DNA sequences and newly introduced DNA sequences.	Defined,predictable phenotypes,no interfering with observation	Requirement-knowledge of targeted gene structure and expression and convenient cassette for each allele *Rare homologous recombination
RNAi	A process in which double stranded RNA direct the sequence spesific degradation of mRNA and knockdown phemotype	Time costless,complement to other genetic studies	Not useful for studying every genes because of in some physiological conditions RNAi machinery is not expressed in some cells and tissues

Table1 Basic approaches for producing genetic change in the mouse

1.2 Gene trapping through targeting

Since the completion of the mouse genome sequence, one major resource to appear has been bacterial artificial chromosome (BACs) libraries. This has now shifted the focus of the field to functional genomics. Modification of the mouse genome through insertion of mutations is central to this field. The previous section gave general information about the principles of widely used mutagenesis approaches, such as ENU induced chemical mutagenesis, and considered their strengths and weakness. Among these approaches, gene trapping and gene targeting play pivotal roles in creating mouse mutations, and are most promising to characterize gene function, structure, and expression.

To produce complete loss of function mutations, gene trap mutagenesis involving plasmids or viral vectors, has up until now been the technology of choice. In fact, after the first release of the mouse genome, gene trapping was the primary strategy to identify mouse genes that had not yet been isolated and cloned. Unlike gene trapping, gene targeting involves homologous recombination between DNA in the chromosome and newly introduced DNA. Depending on vector design, allelic series can be generated.

Both trapping and targeting technologies can be implemented in cultured mammalian and embryonic stem (ES) cells, and injection of genetically modified mouse ES cells into blastocysts can create genetically modified mice. Nevertheless, both technologies have some limitations. Gene trapping can only be employ for certain genes expressed in ES cells. For genes that are not mutable with the gene trapping approach, gene targeting can be a reliable mutagenesis source. Another bottleneck of gene trapping is generating mutant mouse lines **(97,115).** Low rates of germline transmission with trapped ES cell lines give rise to the more sophisticated targeted mutagenesis strategy. Unfortunately there two notable disadvantages of gene targeting: (1) low targeting frequency in mammalian systems compared to other eukaryotic systems such as yeast, and (2) time-consuming protocols. The latter can be attributed to the need to gather information about gene structure and the process of designing a suitable vector for each allele.

The future of functional genomics depends on strategies that employ both trapping and targeting. For this aim it is important to generate a novel vector that will be applicable to the vast majority of mouse genes. This will enable high-throughput generation of mouse mutants.

1.2.1 Gene trap mutagenesis

1.2.1.1 Basic trapping vectors

Enhancer, promoter, and gene trapping vectors are the major vehicles for generating random insertional mutagenesis in mouse ES cells. All three kinds of trapping vectors can be plasmid or viral based and they can include modifications such as reporter genes, e.g., β-galactosidase **(Table 2),** and/or a selection drug resistance marker such as neomycine **(Table 3)**.

An enhancer trap vector is composed of a reporter gene under the control of a minimal promoter and a selectable marker, generally an antibiotic resistance gene suitable for mammalian systems **(Figure 1)**. The expression of the reporter gene in ES

cells is dependent on integration of the enhancer vector into a region near cis-acting regulatory elements that promote reporter gene expression. Because of this, enhancer trap mutagenesis is not feasible to identify transcriptionally inactive genome regions in mouse ES cells. Another impediment to this approach is the difficulty of identifying and isolating a gene by trapping its enhancer, a transcription regulatory region. In complex organisms, enhancers are located at certain distances from the related gene and have the ability to communicate with promoters from this distance. This phenomenon has been shown in the regulation of the homeotic Abdominal B gene (Abd-B) in *Drosophila* **(92)**. Based on this distance model, one of the other limiting factors for enhancer trapping mutagenesis is failure to identify gene by cloning efforts at the vector integration site. Enhcancer traping, nevertheless, has a high trapping frequency and is not limited to only genes that are expressed in undifferentiated ES cells **(50)**. Enhancer traps are especially useful for mapping and identifying transcription regulatory elements **(17)**, despite being less efficient for mutagenesis **(119)**.

Fluorescent based reporter molecules		Enzyme based reporter molecules
GFP(green fluorescent protein) variants	RFP(red fluorescent protein) variants	
EGFP(enhanced green fluorescent protein)	DsRed(Discosoma red fluorescent protein)	Beta-galactosidase
EBFP (enhanced blue fluorescent protein)	HcRed1(Heteractis crispa mutant red fluorescent protein)	Chloramphenicolacetyltransferase(CAT)
ECFP (enhanced cyan fluorescent protein)	AsRed2(Anemonia sulcata red fluorescent protein)	Human alkaline phosphatase(hAPP)
EYFP (enhanced yellow fluorescent protein)	Jred(Anthomedusae jellyfish mutant red fluorescent protein)	Luciferase(luc)

Table 2 Common reporter molecules

A reporter molecule is an easily assayed fluorescent or enzyme based protein encoded by a reporter gene. Among the enzyme based reporter molecules Beta (β) galactosidase is the useful one in mouse ES cell systems. It is highly stable and easy to detect by X-gal staining due to its enzymatic activity. It also has a fusion form, β-geok (β-galactosidase-neomycin phosphotransferase encoded by the fusion gene lacZ-neomycin phospho transferase. Recently green fluorescent protein (GFP) from jellyfish and its mutant variants have been introduced into mammalian cells and transgenic animals. GFP and its variants, such as red fluorescent protein, can be visualized in living cells as well as in early mouse embryos.

Dominant selectable markers			Recessive selectable markers		
Selection marker	Positive selection drug	Negative selection drug	Selection marker	Positive selection drug	Negative selection drug
Neomycine phosphotransferase	G418	-	Hypoxanthine phosphoryl transferase(hprt)	Aminopterin,hypoxanthine,thymidine	6-thioguanine
Puromycine-N-acetyltransferase	Puromycine	-	Xanthine/guanine phosphoribosyl transferas	Aminopterin,hypoxanthine,thymidine	6-thioguanine
Hygromycine B phosphotransferase	Hygromycine B	-	Herpes simplex thymidine kinase(HSVtk)	Aminopterin,hypoxanthine,thymidine	6-thioxanthine
Blasticidin S deaminase	Blasticidin S	-	Thymidine kinase(tk)	Aminopterin,hypoxanthine,thymidine	5-bromodeoxyuridine
Zeocin	Zeocin	-			
Xanthine/guanine phosphoribosyl transferase	Mycophenolic acid	6-Thioxanthine			
Herpes simplex thymidine kinase(HSVtk)	-	Gancyclovir,FIAU			
Dipheria toxin A fragment(DTA)	-	no requirement			

Table 3 Common selectable markers

Selectable markers are reporter protein molecules encoded by antibiotic resistant genes. Because of the low toxicity to mammalian cells and efficient selection features, neomycin is the most commonly used for selection purposes. The minimum requirements for promoters that drive expression of selectable genes are that they should be active in the target cells and that they have no regulatory effect on selectable marker and endogenous gene Expression. But some hybrid promoters like CAGGs (CMV immediate early enhancer fused) have a position effect depending on the genome site of vector integration.

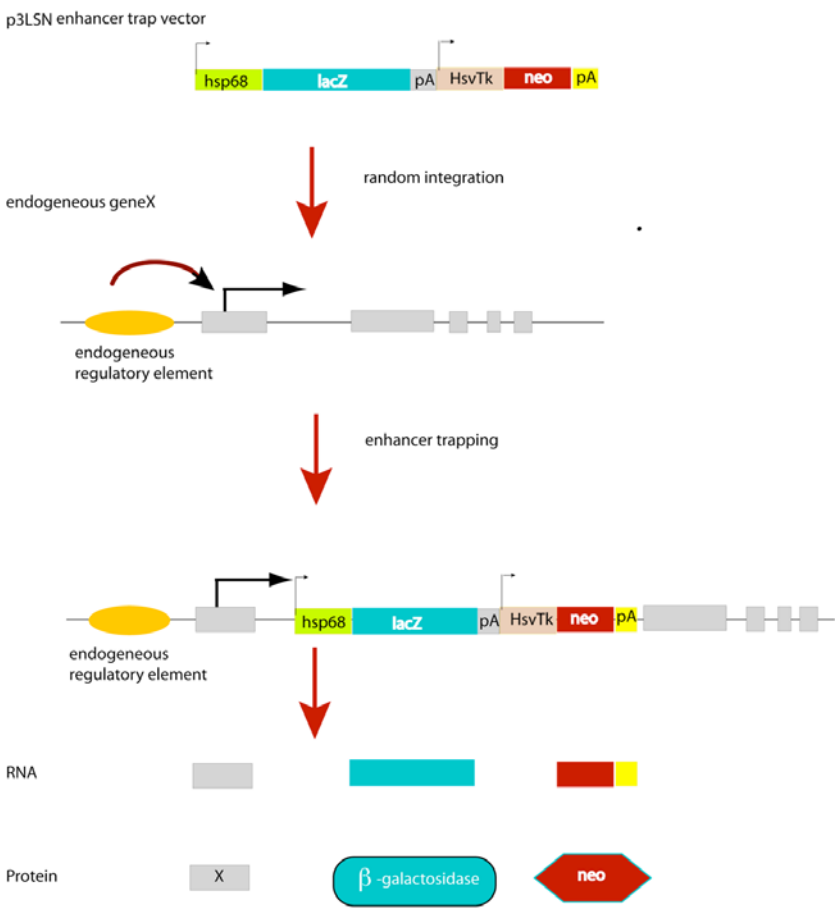


Figure 1 p3LSN enhancer trap vector in action
Enhancer trap vector p3LSN is composed of the hsp68 minimal promoter, lacZ as a reporter marker, and neomycin driven by HSVtk promoter as a selection marker. When the vector randomly integrates into a target gene, X, β-galactosidase and neomycin are expressed. This trapping leads to an incomplete null mutation **(Stanford and Cordes 2001)**.

Promoter trapping vectors were generated to circumvent enhancer vector limitations for gene isolation and mutagenesis efficiency. The general structure of a promoter trapping vector includes a promoterless reporter gene and a selectable marker **(Figure 2)**. Reporter gene expression in this approach depends on insertion of the vector into an exon to generate a fusion transcript of the targeted exonic sequences and the reporter gene. This approach is highly mutagenic, producing loss of function mutations in

ES cells. However, it is restricted to genes that are transcriptionally active in ES cells. It has also very low trapping frequency **(34,47,119)**.

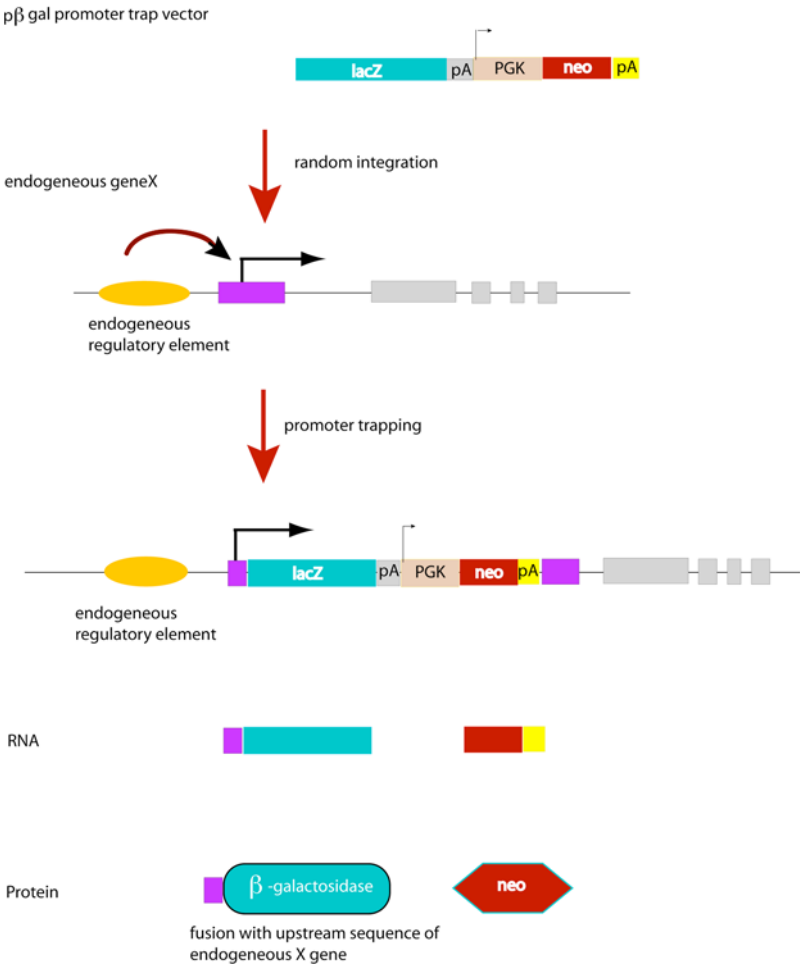


Figure 2 pβ-gal promoter trap vector in action
When the promoter trap vector integrates to the coding sequence of the target gene, X, a fusion protein consisting of upstream endogenous exonic sequence and neomycin is produced. Compared to enhancer trapping, the promoter trapping mutagenesis rate is high, although this is dependent on the strength of polyadenylation element (pA) which truncates the trapped transcript downstream of pA **(Stanford and Cordes 2001)**.

Gene trapping is the most widely used mutagenesis system in ES cells. Two types of vectors can be used depending on the aim: intron and exon trapping vectors. They are characterized by a promoterless reporter gene flanked by a splice acceptor (sA) and a polyadenylation signal (pA). An IRES (Internal Ribosomal Entry Sequence) can also be used in gene trapping vectors for bicistronic mRNA production **(Figure 3)**. A reliable

splice acceptor site (sA) splices to the donor splice site (sD) of the trapped gene, captures the RNA and a pA signal then truncates the transcript. However this depends on the strength of splice acceptor and polyadenylation signal **(114)**. In cases where the sA is weak, the gene trapping vector can be excised, failing to knock-out the gene.

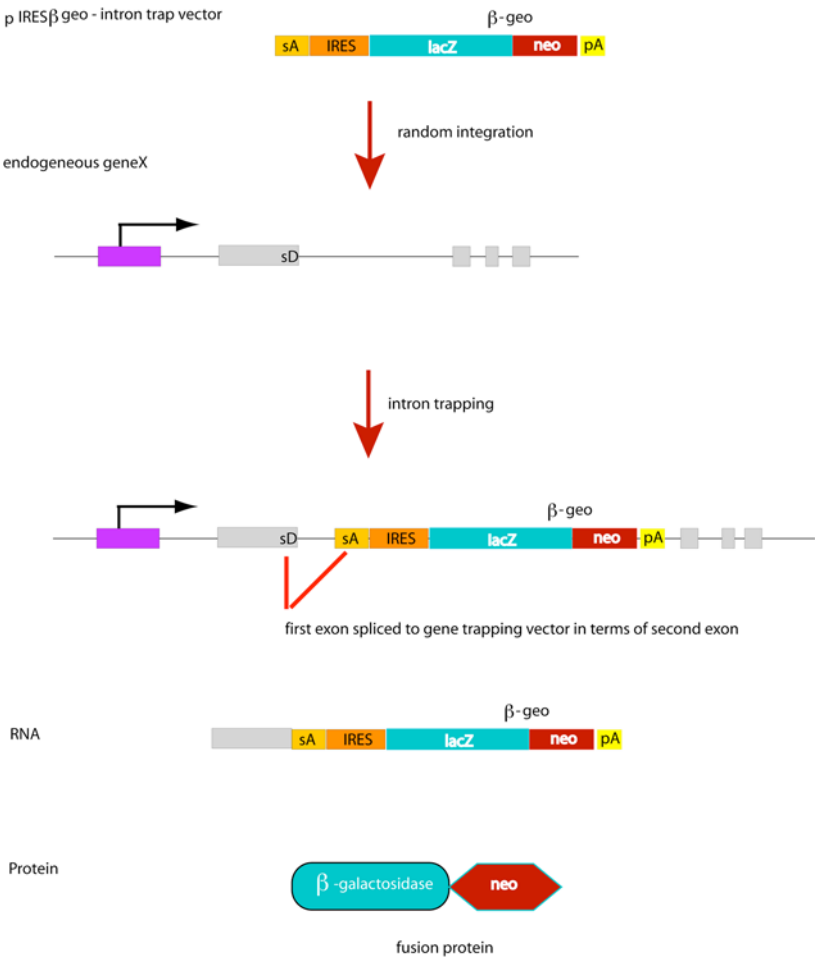


Figure 3 pIRES-βgeo gene trap vector in action
pIRES-βgeo gene trap vector is composed of RNA processing elements. When it integrates to the intron region of target gene, X, it creates a mutation at the RNA transcription level. The sA and pA signals play are critical for the mutagenesis success rate. If the sA signal is weak, splicing machinery will skip the vector and it leads to vector removal. The pA signal should be also strong to prevent downstream gene expression. In the end a fusion protein is produced **(Stanford and Cordes, 2001)**.

Today many public and privately funded efforts are turned towards revolutionizing vectors for efficient gene trapping **(NIH Planning Meeting for a Knockout Mouse Project 2005)**. The improvement of vector design will be one of the critical factors for high throughput insertional mutagenesis in ES cells as discussed in the following chapters.

1.2.2 Gene targeting mutagenesis

1.2.2.1 Targeting vectors

A targeting vector includes: exogenous DNA containing basic elements; sequences homologous to the desired chromosomal integration site; a selectable marker; a reporter gene; and a plasmid backbone. The structural features of the exogenous DNA must be considered to generate the desired mutation at the target region, a high rate of homologous recombination, and efficient identification of recombinant clones. Targeting vectors can be classified as either replacement vectors or insertion vectors.

1.2.2.1.1 Replacement vectors

Basically, a replacement vector is composed of: a 5 to more than 10 kb homology arm for efficient double reciprocal recombination between the vector and endogenous target site; a dominant positive selection marker serving two aims: (i) generation of mutation on exons and/or regulatory sequences of the endogenous target gene, (ii) identification of recombinant cells; a negative selection marker to eliminate various subpopulations that have retained the negative selection marker through the random integration of the vector **(Table 3)**; a linearization site outside of the homology arm; and the plasmid DNA backbone. Target chromosomal sequences are replaced by vector sequences after a double crossing event involving the flanking homologous region **(Figure 4)**.

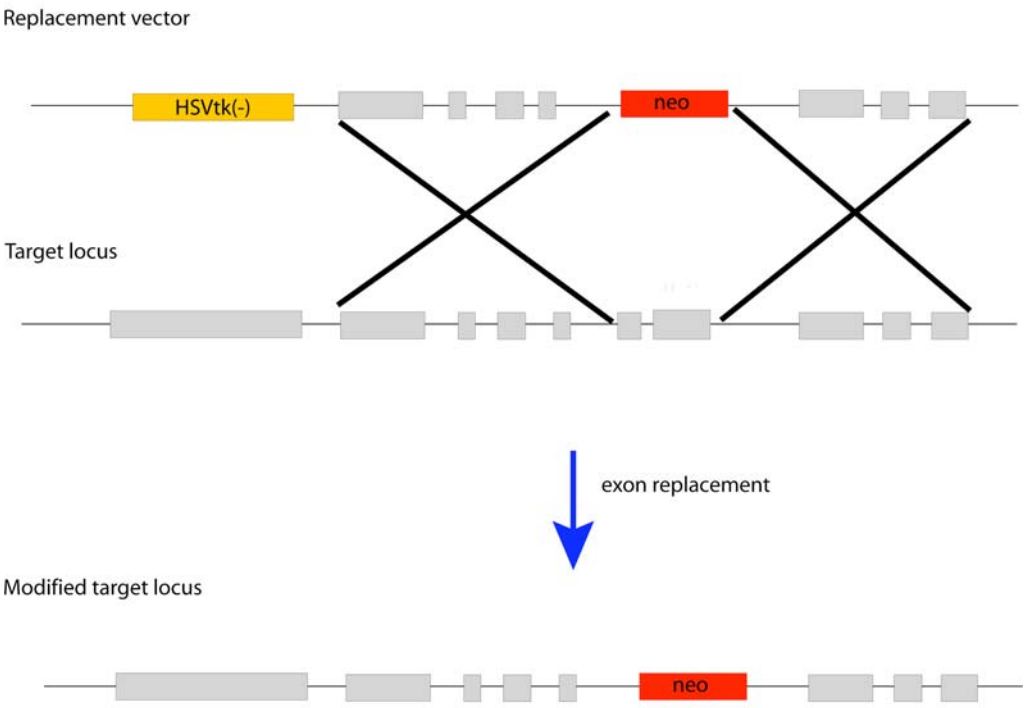


Figure 4 Gene targeting with a replacement vector

In this figure, grey boxes illustrate exon regions of a target gene and crosses show the homologous recombination between the vector and endogenous gene. Following the correct homologous recombination event, the two critical exons of the gene are replaced by the neomycin gene of the linearized replacement vector. In some events, the marker gene only interrupts the exon, instead of replacing it, creating a frame-shift mutation.

1.2.2.1.2 Insertion vectors

Insertion vectors have the same structural features as replacement vectors. The major difference between them is the linearization site of the insertion vectors on the homology arm. Unlike replacement vectors, insertion vectors undergo single reciprocal recombination **(Figure 5)**.

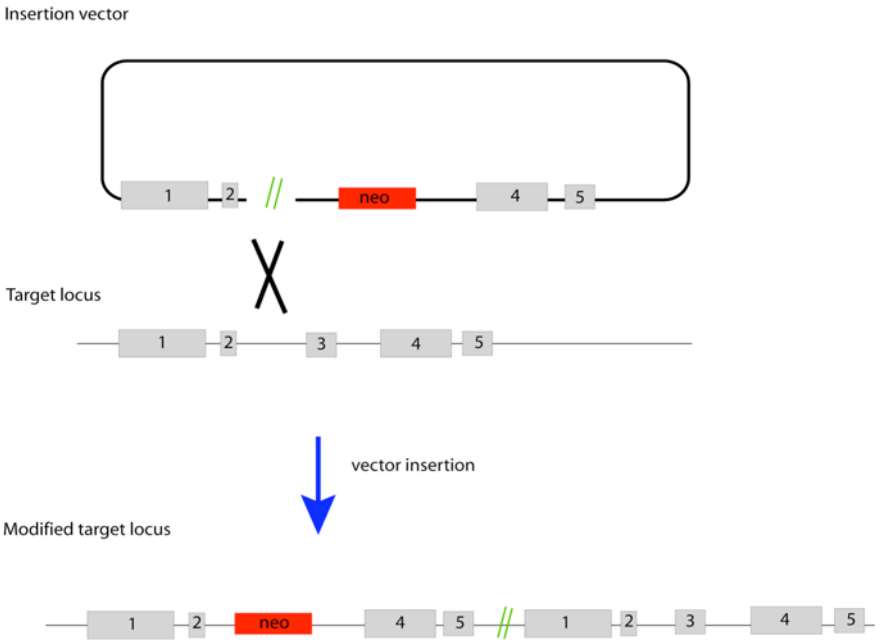


Figure 5 Gene targeting with an insertion vector

The vector linearization site on the homology arm is shown by green double lines. The single reciprocal recombination generates a duplication of the target gene exons, making the insertion vector highly mutagenic.

1.2.2.2 Recombinant alleles generated by replacement and insertion vectors

The first awesome example of mutant allele generation through homologous recombination was the disruption of hprt via insertion of a selectable marker, neomycin into its exonic region **(16)**. Targeted modification of the beta globin locus was done in the same way **(93)**. At that time, every attempt at gene targeting aimed to generate a complete loss of function mutation, namely a knock-out allele. Today the generation of null mutants is also the most common gene targeting strategy. Replacement vectors are

preferred for this strategy. Depending on vector design and targeting strategy, mutations can be generated that yield hypomorphic (partial loss of function), or hypermorphic (gain of function) alleles.

1.2.3 Major topics for effective targeted trapping strategy

1.2.3.1 Plasmid vector design

The development of complex and more versatile plasmid based vectors has been expanded in parallel to the exponential increases in mouse genome information, BAC resources (especially the C57/B6 mouse strain and some 129 substrain genomic DNA libraries), and the unraveling of mammalian biological phenomena such as transcription termination.

Today, both basic trapping and targeting vectors have been revised to generate genetic changes in the genome that vary from single nucleotide mutations to large-scale chromosomal rearrangements. Advanced vector designs combining trapping and targeting vectors have extended their structural elements to handle critical biological processes, e.g transcriptional control. One application for this new recombinant technology is site-directed mutagenesis, a major agenda for many mouse mutagenesis consortiums. Recombineering technology in prokaryotic (*E.coli*) and mammalian systems has made pivotal contributions towards achieving any desired mutation in mouse genome.

1.2.3.2 Contributions to plasmid vector design and construction

1.2.3.2.1 Recombineering technology in *E.coli*

Vector construct assembly is the first step for an effective targeted trapping strategy. Routinely this process consists of screening a genomic library to obtain the gene of interest, followed by restriction mapping and numerous cloning and subcloning steps.

From the desk to the bench, this process seems to be laborious, highly time consuming and very expensive. In addition to this, standard recombination techniques are not feasible for manipulation of large segments. Large constructs are not stable in some cloning vectors. Technology taking advantage of homologous recombination in *E.coli*, allows one to bypass time-intensive, inconvenient cloning steps for vector construction **(28)**.

This strategy has been referred to as recombineering **(20)**, recombination cloning that utilizes *E.coli* strains to enhance recombination between short DNA homology regions, 40-50 bp. Early experimental approaches using the Rac encoded RecET system were invented in our laboratory. The basis of this system is to mediate recombination between short terminal homology stretches (42bp) of PCR derived ds linear DNA fragments and recipient circular plasmid **(123)**. Following successful RecET experiments, it was shown that λ phage mediated recombination events are more efficient than RecET **(66)**. Based on this data, the RecET/Red recombination system was developed in our laboratory. The system components RecE and Redα are $5^{I} - 3^{I}$ exonucleases, and RecT and Redβ are DNA annealing proteins **(66,69)**. The last member, Redγ, serves as a protection factor of unstable linear DNA via inactivation of endogenous *E.coli* host exonuclease RecBCD **(Figure 6)**.

The Red ET recombineering system has been applied to alter various kinds of DNA targets including *E.coli* chromosomes, high copy number plasmids, and BACs. Today the majority of laboratories use this technology to avoid laborious steps such as cloning and subcloning of large DNA molecules (BACs) and to generate 40 kb deletions or insertions **(Figure 7)**.

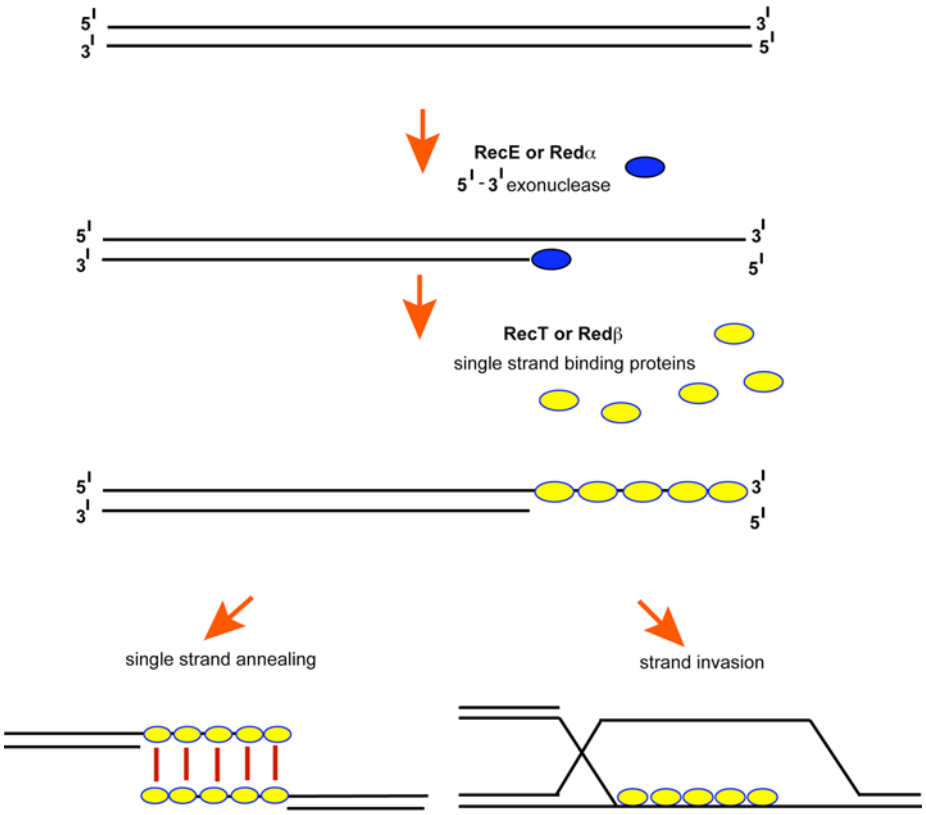


Figure 6 Red-ET recombination mechanism
Red-ET recombination utilizes homologous recombination to repair the double strand break. In *E.coli*, following the double strand break triggered by an endonuclease, RecE or Redα produces two protruding 3^I single-stranded ends (here only one is shown). Single strand binding proteins then bind to single strand DNA. In the strand invasion reaction, 3^I single stranded ends find the homologous region of a second chromosome and make a synapsis. In the single strand invasion reaction, the single-strand DNA binding protein and DNA complex bind to the complementary protein-DNA complex.

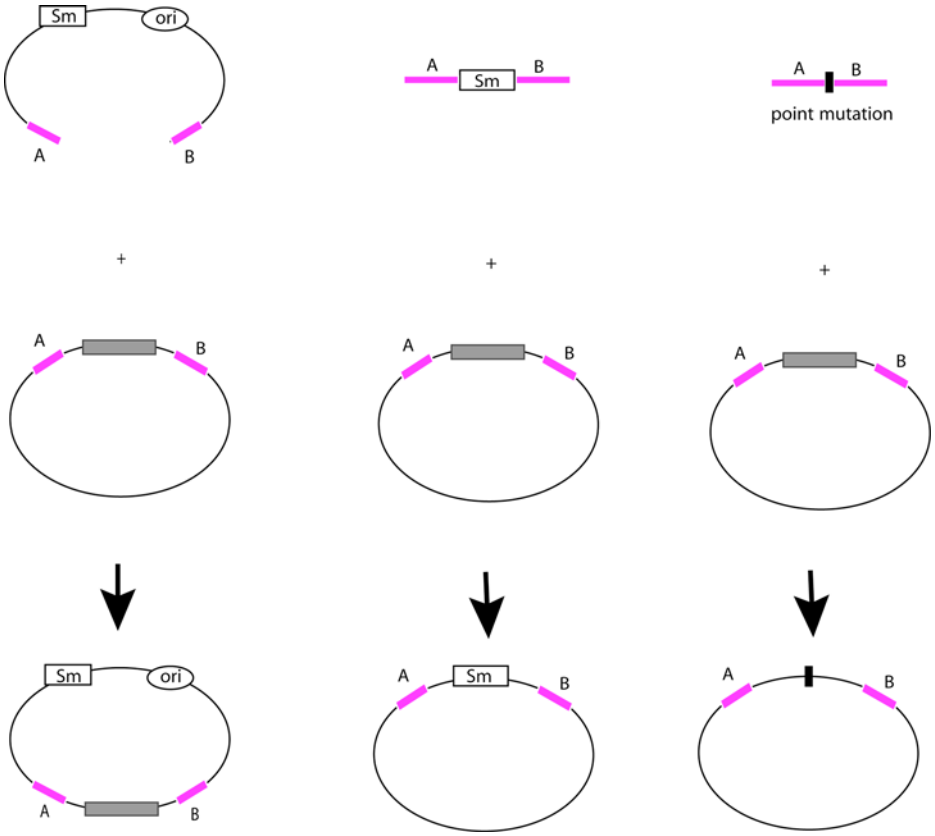


Figure 7 Contributions of Red ET recombination to DNA engineering
Recombination based DNA engineering has a wide variety of applications. The first application (left side) is the subcloning of a DNA sequence from a donor DNA molecule (mostly BAC DNAs for targeting vector generation because it is the primary genomic DNA library source).The linear DNA is PCR amplified plasmid DNA backbone in which the target region from the donor DNA source is subcloned without the PCR requirement that prevents the PCR sourced mutagenesis. The pink lines represent the homology arms. The middle and right figure illustrates the PCR amplified DNA molecules. The middle mechanism leads to selection marker insertion into target DNA. Through this mechanism subtle mutations (right figure) and site directed mutagenesis target sites (e.g loxP) can be inserted.

1.2.3.2.2 Recombineering technology in mammalian system

During the last decade, the development of new genetic tools revolutionized our ability to manipulate mouse ES cells and mice. Earlier reverse genetics studies centered on making complete knock-out mutants. Unfortunately, this research had some critical limitations. First, gene targeting in ES cells requires the use of a selection marker such as neomycin. This marker gene is generally under the control of a strong promoter in ES

cells. However, there is a strong possibility that this promoter will interfere with the endogenous target gene promoter, causing complex phenotypes that are unspecific to the target gene. Second, many genes are developmentally important in mouse. The complete inactivation of a gene can be lethal at early embryonic or post-natal development, preventing the study of its function at later stages in development or in the adult. In order to circumvent these limitations, site-specific recombination systems were developed and added into classical plasmid vectors to allow more sophisticated perturbations to the mouse genome.

20 years ago, engineering the mouse genome through classical gene targeting was a great challenge in molecular biology. But, the refinement of genome engineering using site-specific recombination tools revolutionized this area. Reverse genetics redefined this revolution. From creating small changes, such as deletions and subtle mutations, to large-scale chromosome rearrangements, site-specific recombination is a widely applicable tool in mouse genome engineering.

Cre-lox from the bacteriophage P1 and Flp-Frt from the budding yeast *Saccharomyces cerevisia* are two basic elements of the site-specific recombination tool. Both Cre and Flp are two members of λ integrase superfamily. They share a common DNA recombination mechanism that involves strand cleavage, exchange and ligation at such high fidelity that recombination occurs without any gain or loss of nucleotides and without the need for a cofactor. In the case of Cre recombinase, it has been suggested that the energy required to drive the recombination reaction is derived from the recombinase protein itself.

Despite their mechanistic similarity, Cre and Flp show diversity in terms of DNA recognition sites and temperature-activity. Cre and Flp target sites are composed of 13 bp palindromic sequences, inverted repeats separated by 8 bp asymmetric core sequences, spacers (**Figure 8)**. In addition to the basic target sites loxP for Cre and FRT for FLP, alternative recombinase recognition sites were discovered.

5^{I} ATAACTTCGTATA**ATGTATGC**TATACGAAGTTAT 3^{I}

3^{I}TATTGAAGCATAT**TACATACG** ATATGCTTCAATA 5^{I}

5^{I} GAAGTTCCTATTC**TCTAGAAA**GTATAGGAACTTC 3^{I}

3^{I}CTTCAAGGATAAG**AGATCTTT**CATATCCTTGAAG 5^{I}

Figure 8 Recombination target sites: loxP for Cre and FRT for Flp
The figure shows Cre recombinase target sites, loxP (on top figure), and Flp recombinase target sites, Flp (below). The spacer sequences marked as red nucleotides are illustrated by black arrows on top and grey arrows below. The triangle represents the asymmetry of the spacer for Cre on top, the circle represents the asymmetry of the spacer for Flp. The inverted repeats flank the spacer on both sides.

The recombination reaction involves two highly coordinated strand cleavage and exchange steps. Based on crystallography studies, the integrase family members are known to act as tetrameric complexes. In the first step, two recombinase monomers bind to the inverted sequences on each side of the spacer and create tetrameric complexes together with another two monomers bound to the second recombination target sides. Two opposite recombinases catalyze the first breakage of phosphodiester bond on one strand in the spacer region. Tyrosine 324 residues of both recombinase molecules then make a nucleophilic attack on the DNA backbone, and a specific covalent bond is constructed between the tyrosine and a phosphate entity that is termed a 3^{I}-phosphotyrosine bond. This leads to free 5^{I} OH formation and they attack the partner strand on the other target site to form a Hollyday junction. Before the second strand breakage, the two cleaved strands religate. Then the other reverse recombinase monomers use the same mechanism to promote the second strand cleavage. Following the second two-strand exchange, the final recombination product is released **(Figure 9)**. The relative orientation of these target sites with respect to each other on a DNA segment determines the recombination type including sequence excision, duplication, integration, inversion and chromosomal translocation **(Figure 10)**.

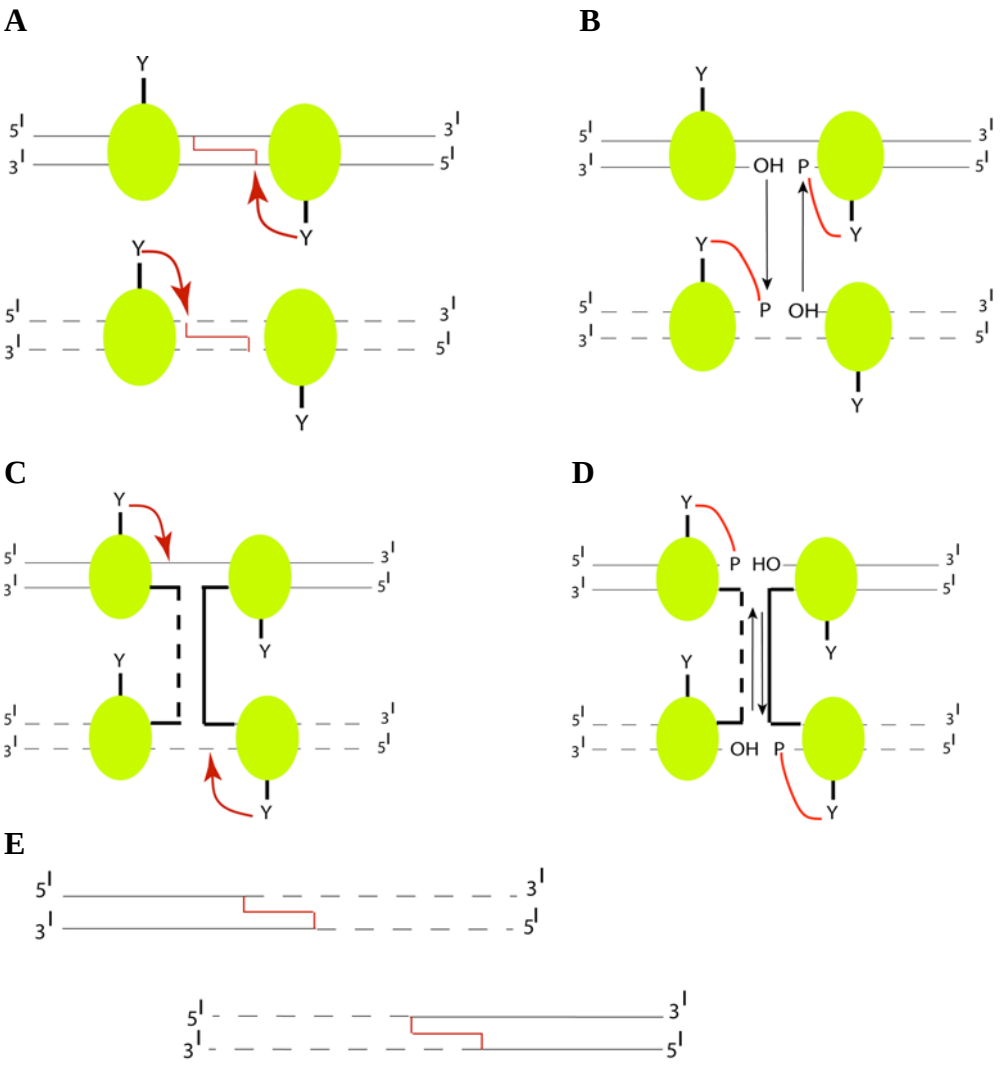


Figure 9 Site-specific recombination mechanism

A) Formation of the tetrameric complex of four recombinase monomers bound to 13bp inverted repeats flanking the 8bp spacer.

B) First step is the breakage of the phosphodiester bond in the spacer region by opposite recombinase monomers. Following this nucleophilic attack, a 3^I phosphotyrosine bond is formed. 5^I OH is then released and attacks the partner strand on the opposite side, thus generating an intermediate holiday junction. The two cleaved strands then religate.

C, D) In the second step, the other recombinase monomer couple undergoes the nucleophilic attack, promoting a similar reaction on the second strand DNA.

E) Completion of strand exchange and release of recombination product.

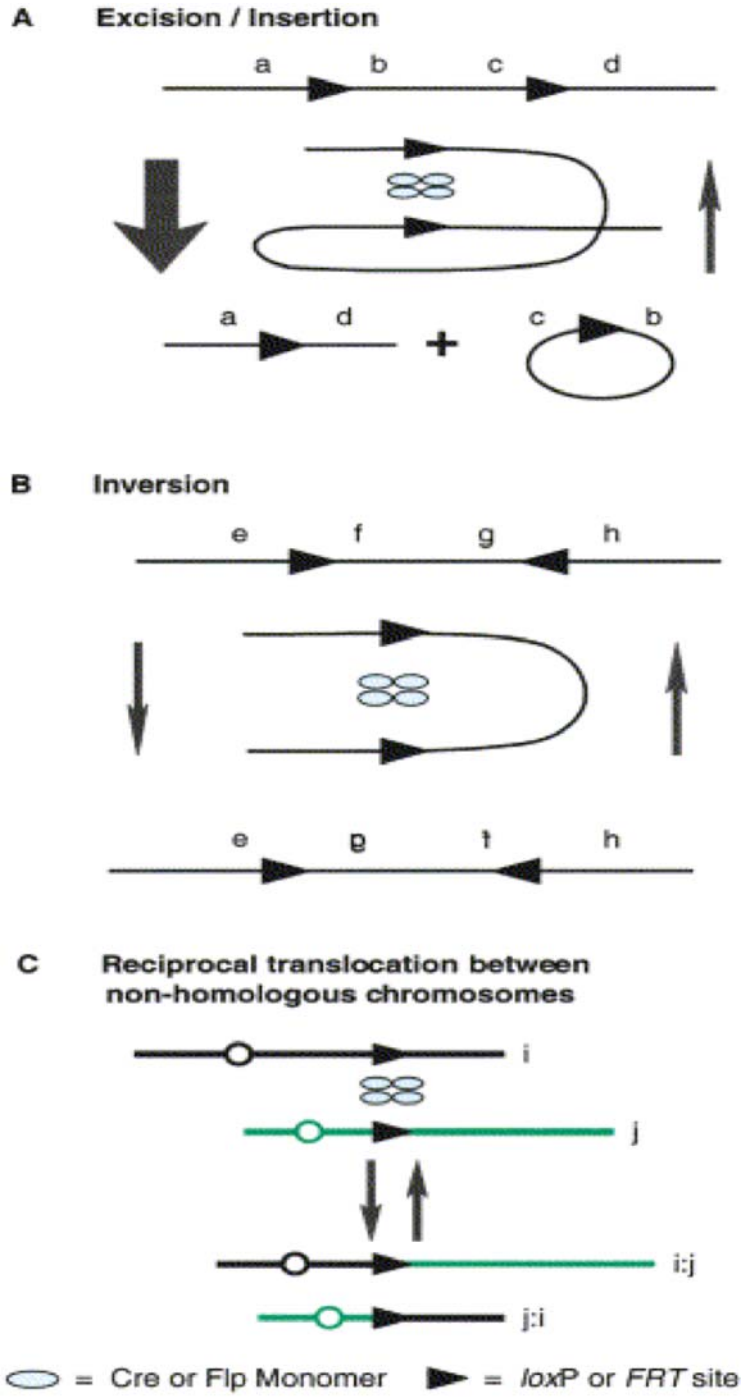


Figure 10 Recombination reactions mediated by Cre of Flp **(Dymecki 2004)**
The relative orientation of target sites with respect to one another determines the recombination types.

A) Excision of DNA segment between the two head to tail positioned recombinase target sites. Products are linear DNA molecule and circular molecule possessing target sites. The reverse reaction is the integration of this circular molecule into the linear molecule.
B) Inversion of DNA segment between two target sites arranged in head to head configuration.
C) Exchange sequences distal to target sites presented on two linear molecules.

Cre and Flp have different temperature sensitivities. As for their activity at 37^0C, Flp is weakl compared to Cre. Cre was shown to be active over a wider range of temperatures than Flp, with maximal performance at 42, making it advantageous for use *in vivo*. To improve the activity of Flp, a mutant variant Flpe (enhanced Flp) was developed by introducing four amino acid changes **(14)**. Flpe has activity equivalent to that of Cre in ES cell cultures.

Combining the Flp and Cre system with homologous recombination will allow an investigator to target cells that would be otherwise inaccessible. The critical considerations of high technology vector design combined with site-specific recombination systems and use genome engineering will be detailed in section 2.3.4.

1.2.3.3 The schemes of vector design for generation of null alleles

Null mutation is a mutation that leads to a complete loss of function of both alleles (autosomal genes and X chromosome-linked genes in females), or a single allele (X chromosome- and Y chromosome-linked genes in males) of the target gene. Basic strategies that can be employed to generate loss-of-function mutations include: (1) deletion of the entire target gene or important gene region, such as the promoter or translation start codon (ATG) by replacement with selectable marker genes, such as lacZ or GFP, or a target gene (cDNA) homologue under the control of cis-acting regulatory elements; and (2) interruption of the coding exon via insertion of a drug resistance marker. Two main types of vectors can be used for these aims in mammalian cells, namely replacement and insertion vectors.

Deletion of an entire target gene by introducing a drug resistance or reporter gene with the replacement vector is a perfect strategy to creation complete null mutations**.** However this strategy is not applicable to the majority of mouse genes because of their size. Replacement type vectors can be applied to generate deletions from several kb up to 20 kb **(62,64)**. The deletion of the entire16 kb mouse Dspp gene, which is responsible for dentin mineralization, with the lacZ-neo-tk targeting construct is strong evidence that this

kind of knock-out strategy results in phenotypes without secondary unspecific effects **(96)**.

Deletions encompassing the promoter of a target gene may inhibit gene expression at the level of RNA transcription. For example, a 2.5 kb deletion in the promoter region of the amyloid precursor-like gene, APLP1, using a PGK-neo-UMS targeting cassette, totally eliminated APLP1 expression **(64)**. This strategy requires strict preliminary study of the promoter and other regulatory elements of the target gene **(42)**.

Another critical point for knock-out allele generation is the selection of exon sites. Null alleles can be created in most cases by introducing a selectable marker into the 5^{I} exons rather than exons encoding the C terminus of a protein **(41)**. For example, targeting the mouse ND gene (a neurological disease gene, Norrie) in the second exon containing the ATG translational codon with a neo-HSvTK targeting vector totally ablated the gene **(8)**. However, not every experimental design based on insertion of a selection marker into the critical exon region of a target gene will create straightforward knock-outs. There are some critical pitfalls that prevent this approach from applying to every gene. One of them is that the mutant exon carrying the selection marker may not be recognized by splicing machinery and skipped. This results in transcripts initiated from the endogenous promoter that may delete the mutant exon or even additional exons from the mRNA. Deletion of the mutant exon or a group of exons with a unit number of codons may result in low levels of protein production. For example, during glutamate transporter gene (GLT1) targeting, the detection of transcript from the mutant GLT1 allele attributed to the lack of exon 4 from the mutant allele **(99)**. In addition to this phenomenon, the excision of the mutant exon or exon groups in some conditions results in modified protein production, mostly because of aberrant splicing. This can be illustrated by the following example. Targeted interruption of exon 8 of the mouse L1 gene, which is responsible for normal nervous system development, by neo-tk plasmid led to a 190 kb mutant isoform of the wild type L1 adhesion molecule in the mutant mouse brain due to exon 8 skipping **(22)**.

The possible neighborhood phenotypic effect in generation of knock-out alleles can be explained by selection cassette interference. Many researchers have shown that the

introduction of a PGK-neo cassette into the target locus causes unpredictable, complex phenotypes, far from simple knock out phenotypes. The targeted inactivation of the mouse myogenic regulatory gene MRF4 carried out by three independent groups is one example of this kind of effect. Three different alleles of MRF4 were generated by Olson, Arnold and Wold by deleting different coding sequences with a PGK-neo cassette that varied in transcription direction. These alleles showed different phenotypes, from lethality, the worst case, to viability. Transcription of the cassette in the same direction as neighbor gene Myf5, another myogenic factor located around 8 kb downstream of MRF4, interfered with Myf5 expression and subsequently decrease its expression **(78)**. This could be explained by the position effect of the PGK promoter as well as the removal of regulatory elements governing the expression of genes within a close proximity. This mostly applies to gene clusters. Thus in addition to avoidance of strong promoter usage, such as PGK in a vector design, strict considerations of the gene structure and organization within the genome, as well as the regulatory element location, cannot be ignored when creating knock-outs.

Insertion vectors, although rarely used for gene targeting, can also be used to generate knock-out alleles. Like replacement vector design, creation of an insertion vector includes some components that promote high-throughput mutation. The basic design steps for insertional vectors are: (1) construction of a homology region to the target locus which has unique linearization site, (2) placement of the exon of the targeting gene on the homology region, and (3) positioning a selectable marker into the plasmid backbone or homologous region. It should be noted that the recombinant allele is characterized by the duplication of the homologous arm and the exon of the endogeneous gene following total vector integration. One very critical point for vector design is to consider possible unintended recombinant alleles generated after vector insertion by a single reciprocal recombination. In most alleles, exon duplications are highly mutagenic. But this depends on some factors. If the duplicated exon does not create a frame shift mutation, the protein will be functional. In order to ensure that the exon duplication was highly mutagenic, stop codons can be introduced into the single exon in the targeting vector. Like in the case of replacement vector null mutations, internal exon mutations

rather than 5’ exons generate incomplete knockouts because the N terminal protein is left partially or totally functional.

Gene targeting with a conventional replacement vector has two limits. In addition to being a low-throughput process compared to random mutagenesis, it requires construction of unique targeting vectors for each new allele. In order to circumvent this limitation and to increase the efficiency of targeted knock-out generation and chromosomal deletions, two 129/Sv (129S5/Sv/Ev/Brd) libraries genomic libraries were built **(124)**. Any clones isolated from these libraries are essentially ready to use vectors composed of genetic elements for gene targeting: a drug resistance gene; the 5^I or 3^I portion of mini hprt gene; a site directed mutagenesis tool; a lox site for chromosome engineering; and a coat-color tag for stock maintenance of mutant mice.

Newly developed insertion vectors, namely 5^I and 3^I Hprt insertional vectors (MICER vectors), carrying sequences homologous to the target, integrate into the specific target region. The recombinant allele generated by this vector is recognized as the duplicated target region separated by vector backbone. This can make them more mutagenic than replacement vectors. The plasmid backbone of MICER vectors has a splicing acceptor to affect splicing machinery and prevent mutant target region splicing. This is advantageous compared to the conventional replacement vector. In addition to their probable higher mutagenecity frequency, MICER vectors have a higher targeting frequency than conventional replacement vectors. These vectors destroy both alleles of a target gene because of different drug resistance markers in trans. They are multifunctional, creating both knock-out alleles and subtle mutations such as deletions and duplications. All these properties make MICER a convenient option to classical replacement vectors **(Figure 11)**.

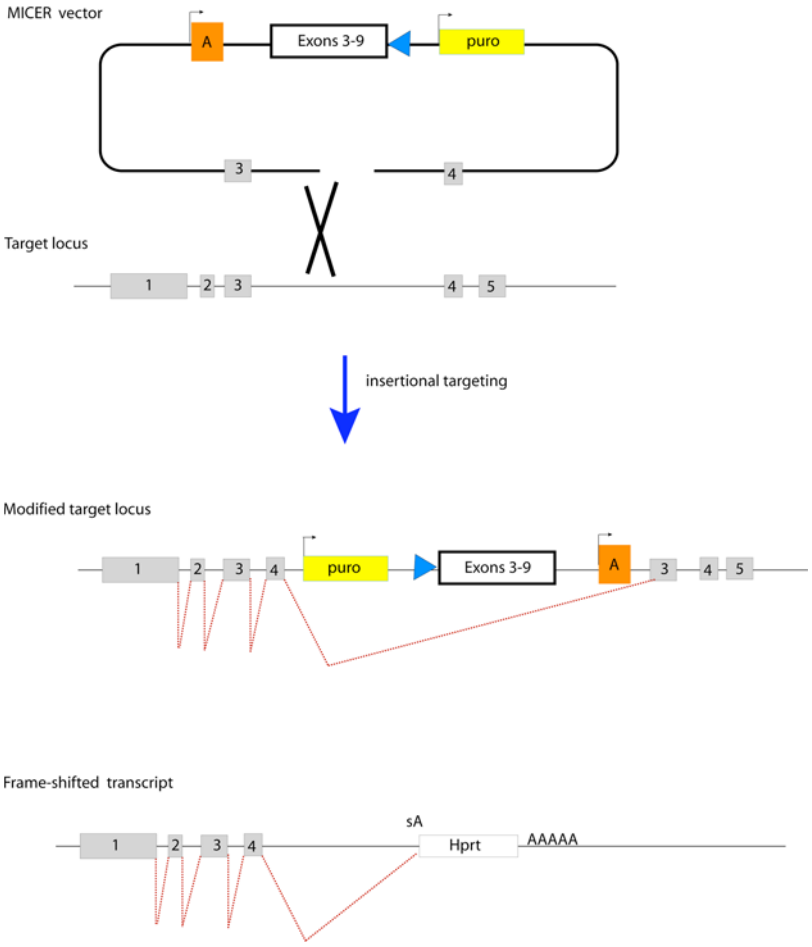


Figure 11 Gene targeting with 3^{I} insertional vector (MICER)
Illustration of gene targeting with a 3^{I} hprt insertional vector which is composed of a coat tagging minigene A (agouti), a lox site flanked hprt exon3-9 fragment and a puromycin resistant gene positioned in the vector backbone. The homologous arm of the vector is composed of single unique enzyme for vector linearization and a homolog sites including exon3 and 4 from the target gene. When the vector undergoes single reciprocal recombination, it leads to duplication of homology arm including exon3 and 4 of the endogenous gene. Exon4 spliced into duplicated exon3 creates a frame-shift mutation that terminates RNA transcription.

An important factor affecting correct gene targeting frequency in mammalian systems is the length of homology to the targeted locus and homology rate. As a result of this, the last consideration for both replacement and insertion vectors design strategies should be construction of the homology arms. The longer the homologous arms, the higher the targeting efficiency, is the key rule for successful homologous recombination in ES cells and mammalian cell types. Therefore, homology arms for both replacement

and insertion vectors should be at least 5, but optimally more than 10 kb, and derived from the same genomic DNA origin (isogenic DNA).

1.2.3.4 Vector design schemes to generate alleles carrying subtle mutations

Simple inactivation of a gene only allows the analysis of a null phenotype. However a series of changes at the nucleotide level of either a coding or a non-coding regulatory region can be more informative about gene function. These subtle mutations include gain of function and knock-down mutations. Two basic strategies can be applied to generate subtle mutations, "hit and run", and "double replacement" strategies.

1.2.3.4.1 Hit and run

Insertion type vectors are generally used in this strategy. In this strategy, a mutation is introduced in two steps. The first step is homologous recombination using an insertion vector to generate a duplication at the target locus carrying the small mutation and to integrate a positive and negative selection marker. Following this homologous recombination, spontaneous intrachromosomal recombination between the duplicated genomic sequences excises the selection markers along with one complement of the duplication from the target locus **(Figure 12)**.

Intrachromosomal recombination depends on the length and degree of homology between the duplicates **(68).** This event can be stimulated by incorporating a rare cutting endonuclease site, most likely 1Scel, into the plasmid backbone. The first example of this hit and run strategy was the insertion of a subtle mutation carried in the second exon of the Hox2.6 locus **(40)**.

1.2.3.4.2 Double replacement

Replacement vectors are used in this strategy. Like the hit and run strategy, the double replacement strategy depends on two rounds of homologous recombination. In the first round, a positive and negative selectable marker is introduced into the target region.

The second homologous recombination requires gene targeting, unlike the hit and run approach, to introduce a subtle mutation by another replacement vector carrying the mutation. This strategy introduces several mutations simulataneously into the target locus in only one round of recombination **(Figure 13)**.

Compared to the hit and run strategy, the replacement strategy is very useful in studies of gene structure and functionality. Deletion is generated in the very first step and mutations can then be made that individually or in combination target the functional domains of the gene. Briefly, this strategy offers an opportunity to introduce a series of deletions or point mutations into regions of a gene that are responsible for regulating expression. So far this approach has been applied to the ColA1gene **(109)**.

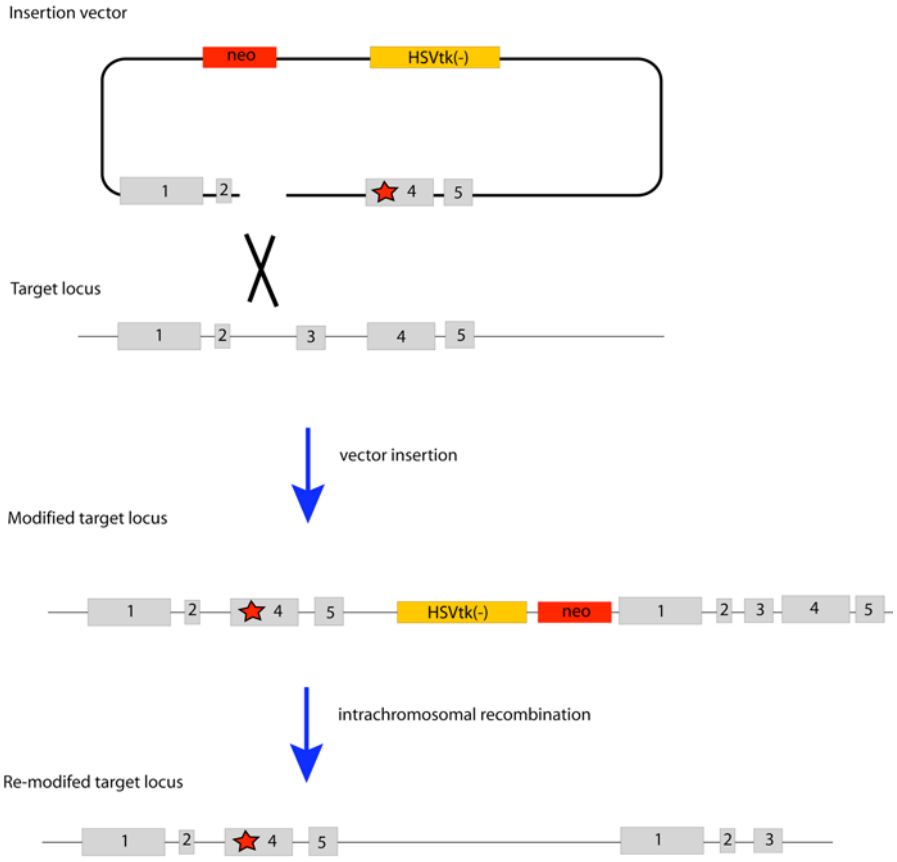


Figure 12 Hit and run strategy

The first step is the insertion of vector to the target region via single reciprocal recombination that leads to duplication of $5^{I}$and 3^{I} endogenous gene regions. The second step includes intrachromosomal recombination or uneven sister chromosome exchange. This leads to the excision of plasmid sequences, selection markers (neomycin as a

positive, HSVTk as a negative selection marker) and one complement of the duplication region from the target locus

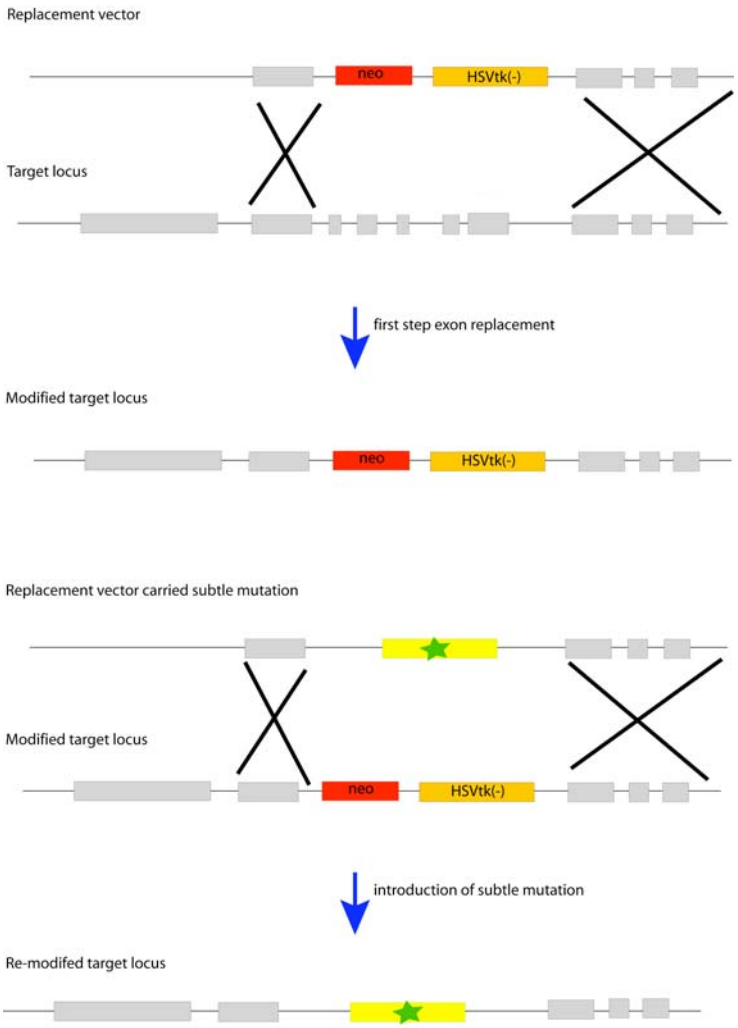


Figure 13 Double replacement strategy
The first step is the deletion of the target region by replacing it with selction markers, neomycin for positive selection and NS HSVTk for negative selection. The second replacement targeting vector carrying the subtle mutation (yellow star) then undergoes homologous recombination and the mutation is inserted into the target region.

1.2.3.5 Expanding the frame for allele generations: High technology vectors meet SSR tools

A wide variety of sophisticated and versatile gene trapping and targeting vectors tmodified to contain site directed recombination tools generate innovative alleles to introduced desired mutations into the germ line. The most popular application of these

vectors is the generation of specific mutations. This strategy requires a selection marker in the targeting vector that is then excised along with the all or a part of the endogenous gene . Basic replacement vectors carrying a positive selectable marker flanked by either loxP or FRT target sites are mostly used for this strategy. For this kind of vector construction, orientation of the site-specific recombinase target sites is critical. Deletions will occur in the interval of DNA between two direct repeats. In the presence of Cre or Flp, the selectable marker cassette is removed. This elegant and time saving strategy prevents position effect of a selection cassette under a strong promoter, thereby circumventing genetic ambiguity arising form changes in neighboring gene expression.

Classically, gene targeting involved in generating null mutations by introducing a selectable marker to the target gene or an essential part of this gene. Deletions of 5^{I} coding regions are a convenient way to create a nonfunctional gene. To the same extent as the deletion sites of endogenous gene, the size of the deletion plays pivotal role in the generation of a complete null mutation. Ideally, complete non-functional mutation requires larger deletions. Nonetheless, the size of the target gene or fragment that can be removed is limited by size when using a classical replacement. Coupled with site-directed mutagenesis tools, more than 30 kb deletions can be created in the mouse genome using classical gene targeting.

Critical points for vectors that combine site-directed mutagenesis tools with classical trapping and targeting vectors include: the relative position of recombinase sites and marker genes; the placement of floxed or flrted gene markers; and whether, following recombination, the loxP or FRT sites will be left to use for future genetic manipulations. To create conditional null mutants, it is essential to position recombinase target sites around the entire gene. Nevertheless, this is only possible for monoexogenic or short genes. Alternatively, an essential part of the target gene can be selected for the construction of recombinase target sites. The target sites should be positioned in a manner to generate total mutation upon the recombinase exposure. A suitable position of the target sites should be selected that does not compromise functionality of endogenous gene or its neighbors, yet keeps the conditional allele fully functional for recombination.

The mechanism of target site recognition and recombination stimulation emplyed by recombinase enzymes has a profound effect on the location of target sites. In biological systems, some of proteins locate their target sites very rapidly, the likely result of diffusional collusions between the protein and DNA molecule. Such mechanisms generally involve the initial binding at a random DNA site via non-specific, sequence-independent interactions, followed by intramolecular translocation to the specific binding site. This seems simple if one ignores the fact that the protein searches megabases of non-target DNA. Hence, recombinases that rely on random collision activity can be affected by the distance to their target sites. Moreover, previous studies that used Cre mediated deletions in somatic cells have pointed out this fact **(125)**. Although, deletions up to 60 cM have been achieved in ES cells *in vitro*, Torres and Kuhn reported that this frequency decreased from 70% over a distance of 4kb to 2% over a distance of 10kb. In summary, the position of a second recombinase target site should be within 10 kb of the selection marker **(Figure 14)**.

Further modifications of the vector can be made to expand the application of the site directed mutagenesis strategies. Introducing subtle mutations and swapping the endogenous sequences for exogenous ones are other applications of these kinds of vectors. General replacement vector configuration serve as a guide for vector construction, with the following exceptions: the small mutation should be included on the homology arm and floxed or flrted; and, a selectable marker should be directed into a large intron or into a region downstream of the pA to avoid putative interference of the single recombination target site following recombinase induction. It is critical to distinguish the wild-type allele and recombinant allele by incorporating a unique restriction enzyme into the small mutation site **(Figure 15)**.

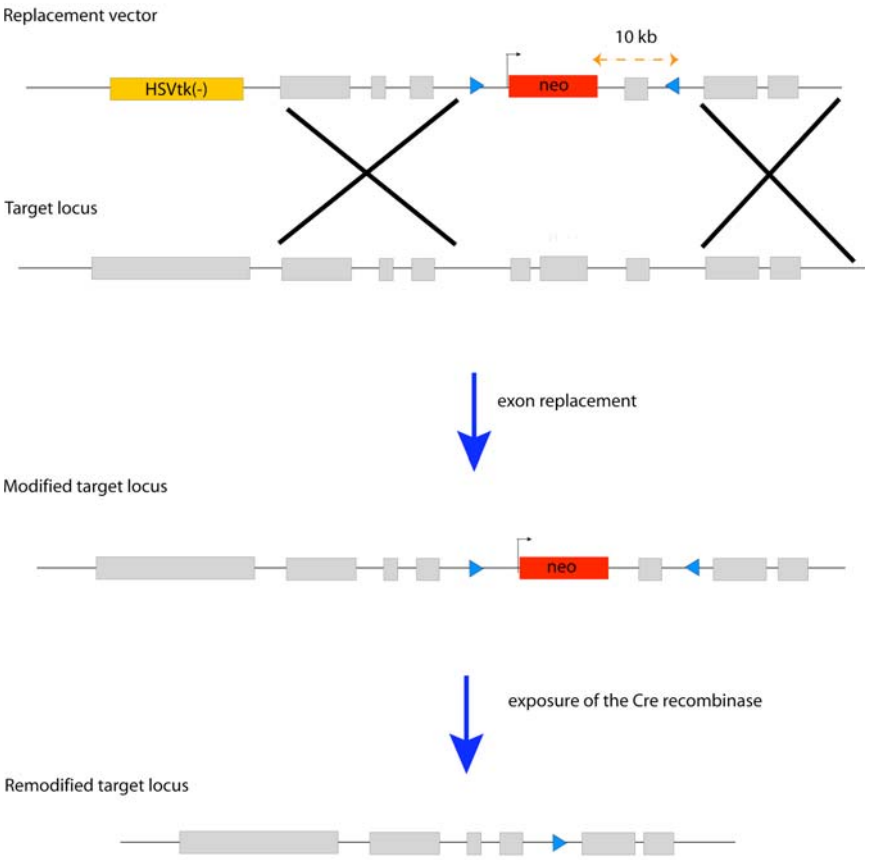


Figure 14 Gene deletion along with the marker removal through site directed mutagenesis

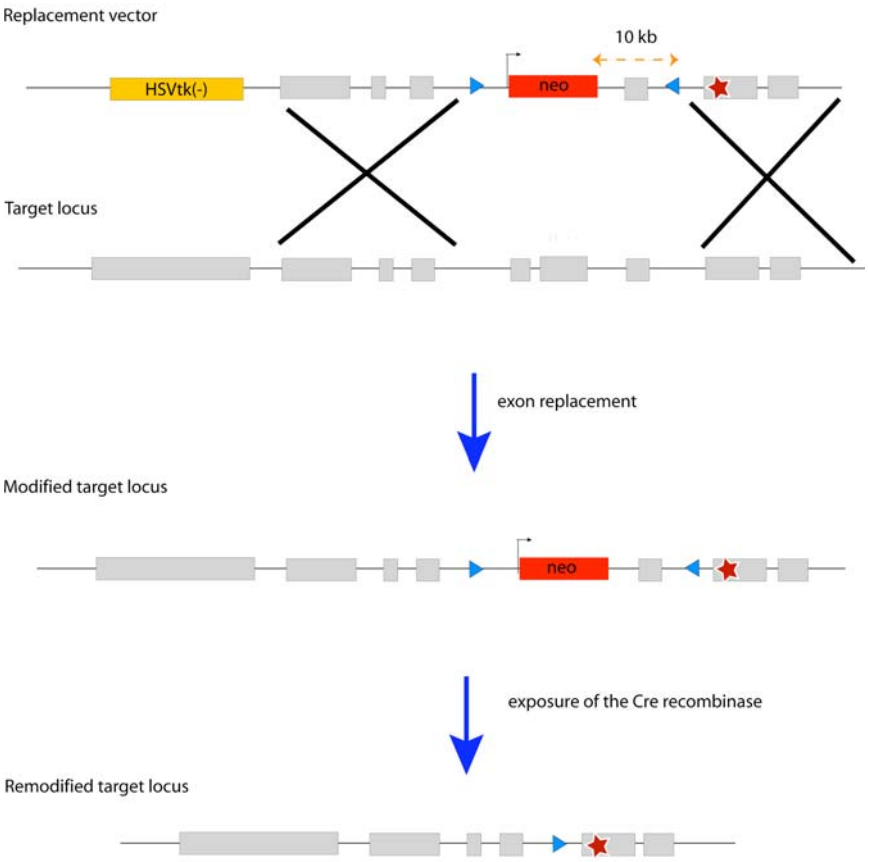


Figure 15 Introduction of a subtle mutation along with marker removal

The figure shows gene targeting with a replacement vector containing a subtle mutation in one of the endogenous exons and two lox sites in opposite orientation. Following the insertion of the mutation, Cre recombinase is induced and this removes the positive selection marker, neomycin.

Another vector configuration enables the creation of knock-in alleles. Although the vector configuration pattern is the same previous ones, the design approach includes the incorporation of exogeneous sequence into the selection marker gene **(Figure 16)**. This generates heterology in one homology arm of the vectors, but as some studies have shown, this heterogenecity does not affect homologous recombination.

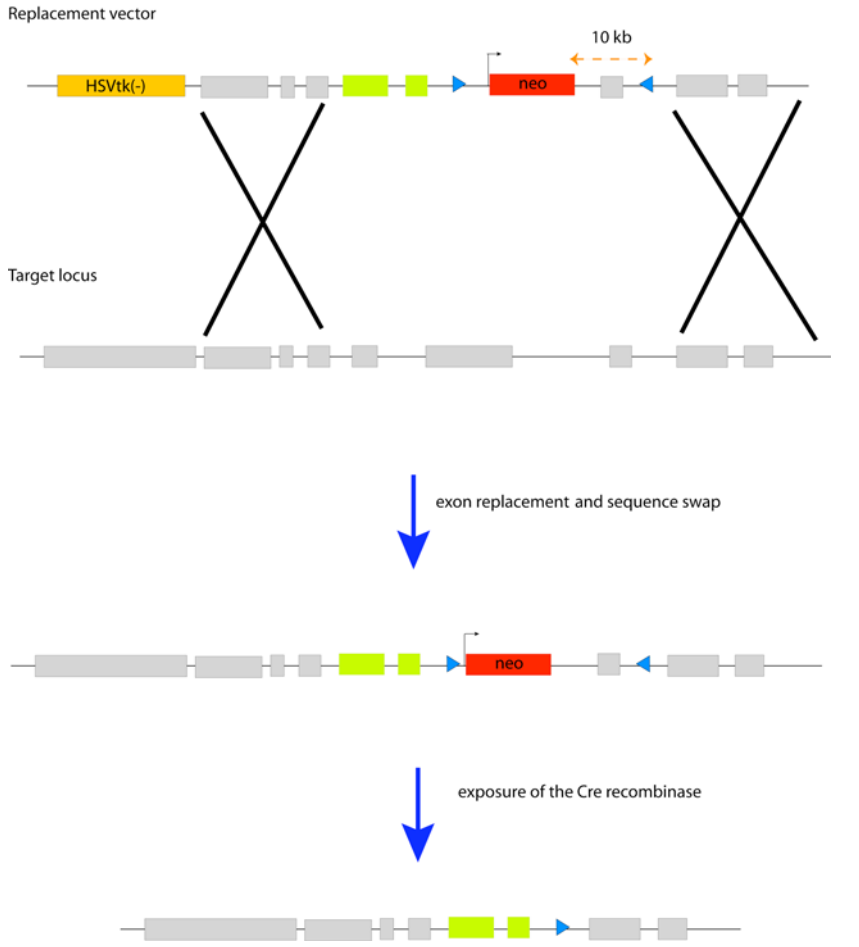


Figure 16 Sequence swapping along with exon replacement and marker removal
Gene targeting with a replacement vector that includes an exogenous sequence as a heterogeneous site in the left homology arm (green rectangle), and a positive marker flanked by two lox sites positioned in opposite orientations. Insertion of the exogenous sequence via homologous recombination is followed by site directed recombination and removal of the positive selection marker neomycin.

Alternative to wild type, the usage of mutant site directed target sites make creation of more functional and elaborate allele designs possible. For example, the placement of two pairs of heterotypic mutant lox sites, lox51 and lox71, around the essential part of a target gene creates inversion not excision **(Figure 17)**. The flexibility obtained by combining two site directed mutagenesis systems, Cre and Flp was an innovative step to generate multifunctional alleles.

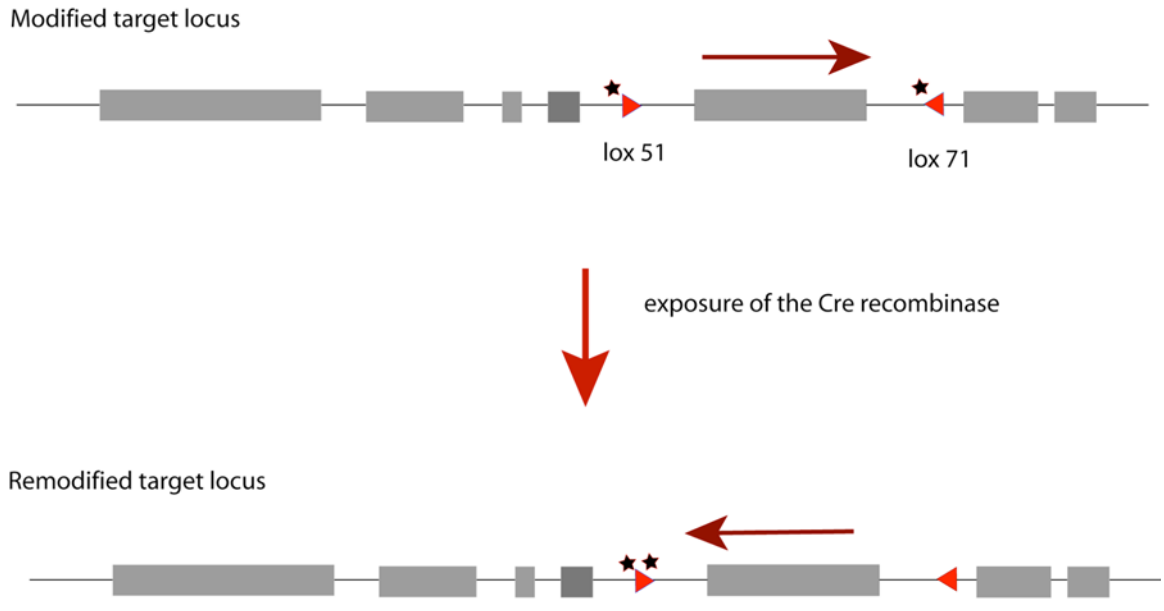


Figure 17 Inversion

The inversion reaction between two mutated loxP sites, lox 51 and lox 71 (stars). Following Cre exposure, the DNA fragment between the mutant lox sites is inverted and a double mutant lox is generated which cannot be recognized by Cre recombinase.

1.2.3.6 Multipurpose Allele Strategy - Cornerstone of the targeted trapping approach

Gene trapping has liberated scientists from laborious genetic screens and biochemical approaches to identify new genes and molecules involved in highly complex biological events in mammalian systems. Gene trapping provides a powerful and rapid approach for simultaneously mutating and identifying genes. Scientific groups in major American and European centers have made remarkable progress in generating ES cell libraries by making large-scale knock-outs of mouse genes using the gene trapping approach. In accordance with the International Gene Trapping Consortium declaration in 2006, gene trapping efforts have resulted in approximately 60% of mouse genes being trapped and the public availability of ES cells. Though gene trapping can be termed a high throughput, large-scale mutagenesis method, it has limitations detailed in section 2. On the other hand, the alternative approach, a direct strategy based on homologous

recombination, has some remarkable pitfalls related to recombination frequency in mammalian systems.

Incorporation of gene trapping-like cassettes and site-specific mutagenesis tools into these basic vectors was a powerful invention to both create mutations and identify new mouse genes. Instead of generating different allelic variation one-by-one, these innovative vectors allow multiple allelic variations to be created simultaneously. Also, it circumvents embryonic lethality and permits the dissection of mammalian gene function and regulation at all developmental stages and in adulthood.

1.2.3.6.1 Multipurpose strategy to create knock-out alleles

This strategy relies on the knock-out first approach in combination with conditional applications. An advanced replacement vector is used for this multifunctional allele creation. This vector incorporates gene trapping cassette that consists of efficient RNA processing signals (sA and pA), a promoterless reporter gene (lacZ, GFP or RFP), and a positive selection marker such as a drug resistance gene (neomycine or hygromycine). This cassette is flanked by a site directed mutagenesis tool, the Cre-lox system. As an alternative to the Cre-lox system, the Flpe/FRT system can be utilized for conditional knock-out strategies in ES cells with the same efficiency **(14)**.

The principle of this approach is like the random gene trapping approach, the vector inserts in frame into an intron region of a mouse gene and traps the upstream exon regions of endogenous gene. One critical consideration for vector design is the RNA processing signal strength. One of the pitfalls is poor splicing of the gene trap cassette because of a weak sA signal. An efficient pA signal in this vector truncates the transcript so that endogenous gene is not transcribed into mRNA downstream of the cassette site. Because the reporter gene is driven under the endogenous gene promoter, this vector reports the promoter function of endogenous gene. Under some conditions, IRES also can be inserted into the trap cassette to produce bicistronic mRNA. An important consideration for vector design is that when there is significant distance between the lox sites, the efficiency of Cre recombinase is diminished. Therefore lox sites should be

placed as close possible to each other in the vector. The homology arm of this chimeric vector for efficient recombination should be at least 5-8 kb or optimally 10kb in each arm.

A successful lacZ reporter gene trap style cassette for a multipurpose knock-out first allele was constructed in our laboratory and used to generate a multifunctional Mll2 allele. The strategy to generate the knock-out Mll2 allele was based on inserting a gene trapping cassette in an intron of Mll2, producing a knock-out at the RNA processing level. In the transcribed pre-mRNA, exon 1 is spliced to the cassette instead of exon 2 because the strong sA, engrailed 2, in the cassette captures the RNA transcript **(35)**. The cassette consists of IRES **(48)**, reporter gene lacZ, and the selection marker neomycin **(34)**. It permits the expression of this fusion protein driven by the Mll2 promoter. The strong pA, SV40 pA, in the cassette truncates the transcript so that the gene can not be transcribed into mRNA downstream of the cassette site. This gene trapping style cassette is flanked by FRT regions that allow Flp to excise the cassette from the intronic region of Mll2 to generate a conditional allele. By introducing another site directed mutagenesis site to the second exon of the gene, this exon can be removed and a frame shift mutation generated by the splicing of exon 1 to exon 3. This frame shifted mRNA includes some stop codons. This converts the conditional Mll2 allele to a constitutive null **(Figure 18)**.

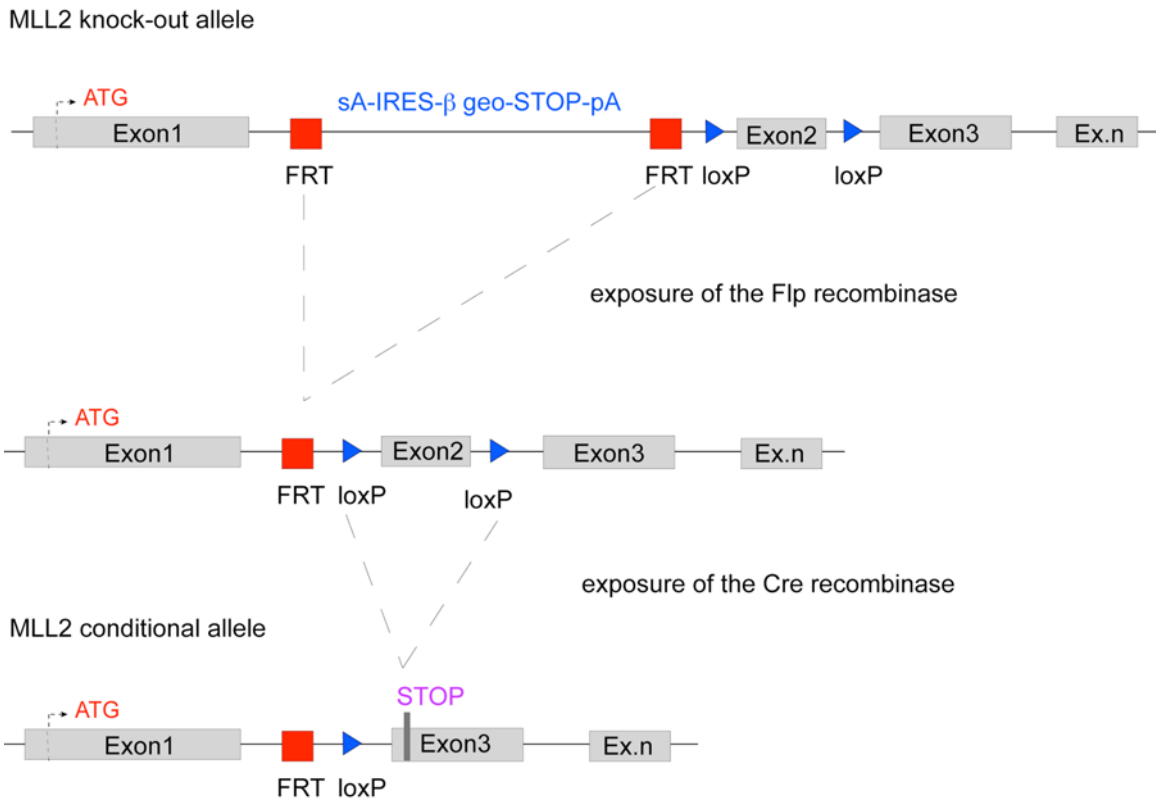


Figure 18 Knock-out first multipurpose allele strategy

The Mll2 multipurpose knock-out allele vector is generated by a gene trap cassette that includes: a splicing acceptor, IRES, a lacZ-neomycin fusion gene, β-geo, a stop codon and a SV40 pA signal that are flanked by FRT on both sides; loxP sites; and a homologous arm for MLL2 target region. The second loxP site within the homology arm of the vector is inserted into exon 2 as part the selection marker cassette flanked by two loxP sites positioned opposite orientation, that allow removal of the selection marker via Cre induction through Red-ET recombination. The cassette in the first intronic region of the Mll2 gene generates a Mll2 null allele by trapping and truncation of the endogenous mRNA. Only lacZ-neo is expressed. Flp mediated recombination leads to cassette removal and the Mll2 gene is expressed again. The Mll2 conditional allele is converted into knock-out allele via Cre mediated recombination. The deletion of exon 2 creates a frame-shift mutation. Splicing of exon 1 to exon 3 generates a stop codon that truncates the mRNA.

1.2.3.6.2 Multipurpose strategy to create conditional alleles

Conditional mutagenesis relies on either deletion of an early frame-shifting exon or deletion of entire gene if it is small. This can be accomplished by placement of two loxp or FRT sites on either side of an essential section of a gene or entire a gene. Nevertheless, this is impractical unless the gene is very small and/or the exon is frame-shifting **(62,64).** Thus a gene trapping style lacZ-neo cassette for the knock-out first multipurpose allele strategy cannot be applied for many mouse genes. The design of each conditional allele requires careful consideration of gene structure. In this PhD study, another gene trapping style cassette called “generic FLEx” was created that does not need to consider individual gene architecture.

FLEx is a strategy for inversional site pecific recombination with Cre. General inversional site-specific strategies are enhanced by the placement of mutant lox sites, lox66 and lox71, or wild type loxp and mutant lox5171 on either side of a DNA fragment in a tail to tail configuration. The resulting recombination inverts the intervening sequence **(Figure 19)**.

Starting with this stunning strategy, a generic FLEx gene trap style cassette was built. The fundamental elements of this cassette are: a splicing acceptor; a polyadenylation signal; and a selection marker. In a novel application, here, transcription termination signals were included in this cassette. This cassette is flanked by loxp and lox5171 in either side, positioned tail to tail for inversional site specific recombination. The basic approach for generic conditional mutagenesis with this cassette is its placement into an early intron. In one orientation, this cassette should be neutral so that it is not recognized by splicing machinery. Following Cre exposure, the cassette should turn to the mutagenic orientation which captures the splicing of the transcript and then terminates it, thereby preventing mRNA production.

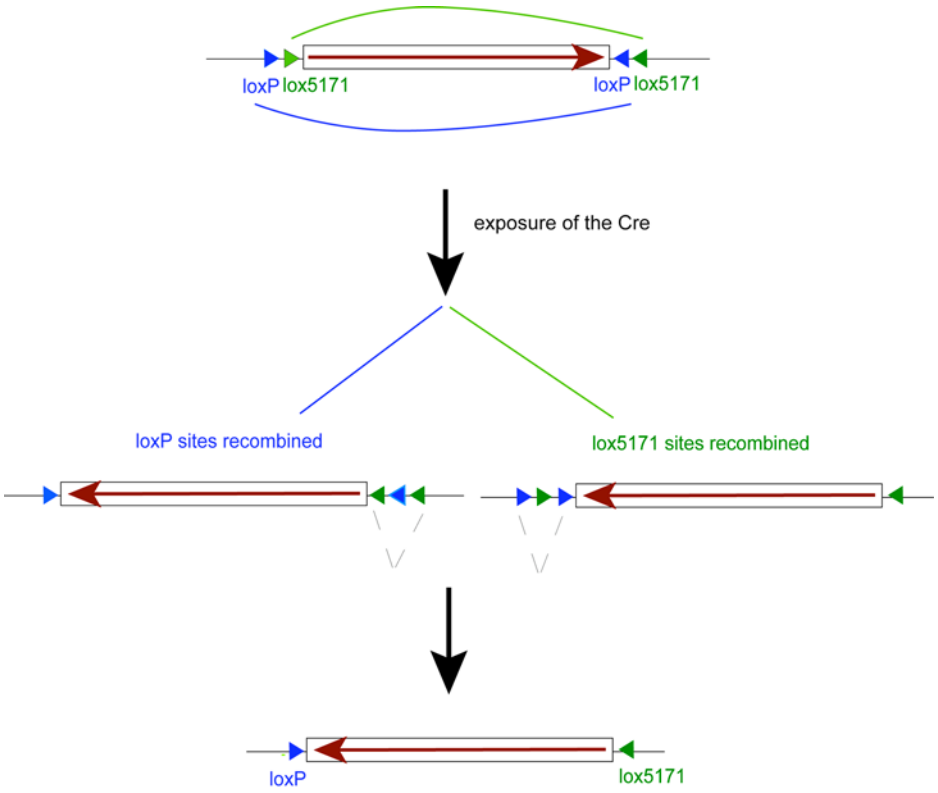


Figure 19 FLEx Strategy
The figure shows a vector cassette combined with a site specific mutagenesis tool. Two pairs of heterotypic lox sites, loxP and lox5171 are positioned in an inverted figuration on each side of the cassette. Upon Cre induction, two homotypic sites, either loxP (represented by blue) or lox5171 (represented by green), are recombined. The cassette inversion leads to uni-direction lox sites that produce a deletion.

1.2.3.6.2.1 The Jarid1c Multipurpose Allele

The application of the targeted trap approach with the FLEx cassette to the X-linked mouse gene Jarid1c is the second critical aim of this project. The strategy to generate a conditional Jarid1c allele is based on inserting a neutrally oriented FLEx cassette into an early intron. In the transcribed pre-mRNA, exon1 is spliced to exon 2 instead of the cassette. Because the splicing machinery cannot recognized the cassette, it skips the oppositely oriented, meaningless, sA and instead recognizes the endogenous sA. Following Cre recombination, a null Jarid1c allele is created. This time the endogenous transcript is trapped by a synthetic splice acceptor. (synt sA). The synthetic polyadenylation signal (synt pA) truncates the trapped transcript **(55)**. Because it is not as strong as other polyadenylation signals, such as SV40 pA, the strong transcription

termination signal CoTC ensures that Jarid1c is not expressed from the downstream region of the cassette. The selectable marker blasticidin is driven by the SV40 promoter. In the mutagenic orientation, this promoter is excised by Cre recombination to prevent interference with endogenous and neighboring gene expression that could create an unexpected phenotype **(Figure 20)**.

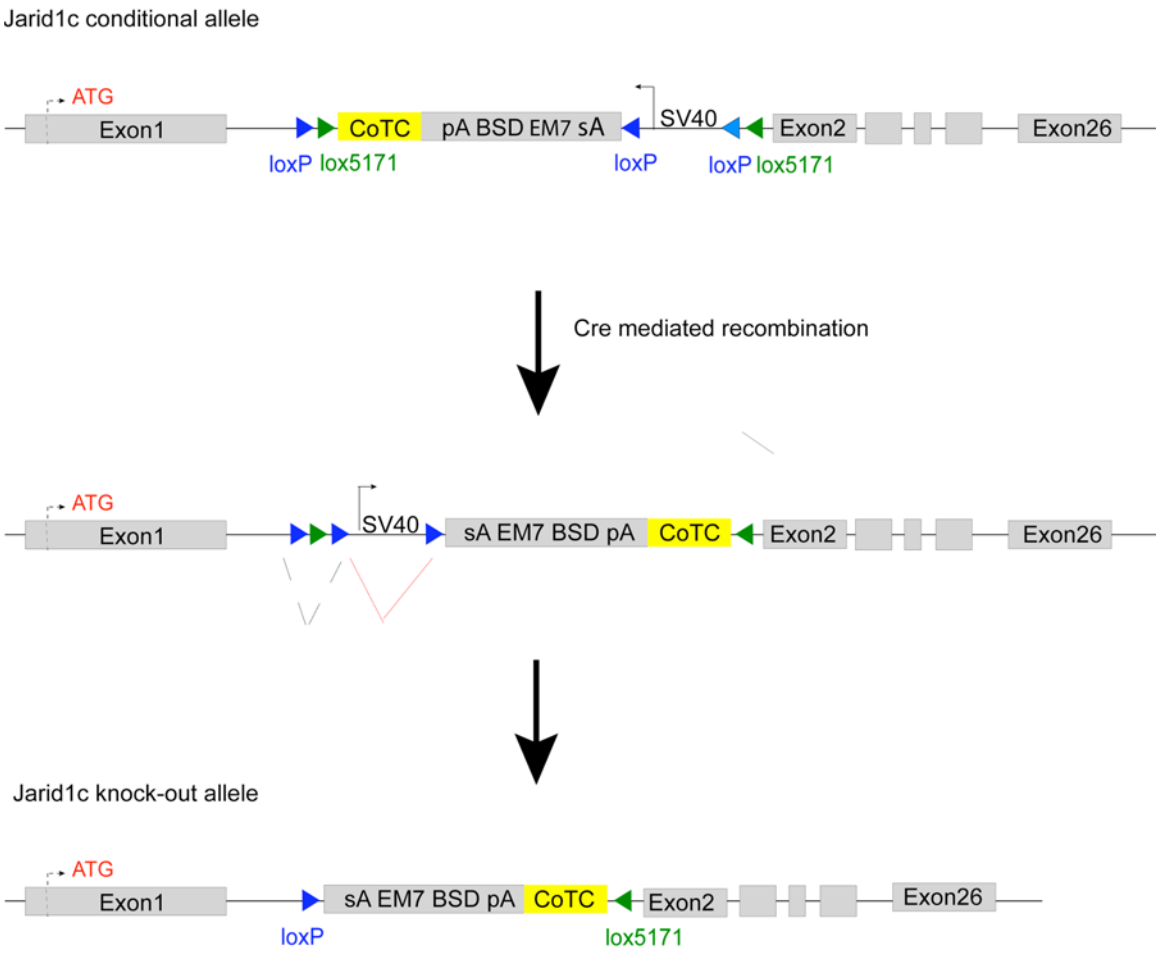


Figure 20 Conditional first multipurpose allele strategy

The conditional allele of Jarid1c. The vector is a FLEx cassette that contains: a synthetic splicing acceptor, the blasticidin (BSD) gene under control of both the bacterial promoter EM7 and the mammalian promoter SV40, the strong transcription termination site CoTC, heterotypic lox sites flank the cassette, and 10 kb left and right homologous arms for Jarid1c excluding its promoter. In the conditional state of the gene, the neutrally oriented FLEx cassette inserts into the first intronic region of the gene and does not interfere with splicing of endogeneous Jarid1c. Exon1 splices to exon 2. By Cre mediated recombination, the cassette inverts into the mutagenic orientation. Splicing machinery skips the endogeneous splice acceptor, exon 1 is then spliced into the cassette instead of exon 2. This prevents Jarid1c gene expression downstream of pA and the transcription termination sites. Transcription termination site CoTC is novelly used inthe trapping style targeting studies. It successfully prevents gene transcription playing a regulatory role in RNApol II dependent transcription termination (detailed in section 3.1)

1.2.3.7 Ideal ES cell and isogenic genomic DNA sources

ES cell lines are cultures of pluripotent, unspecialized cells derived from epiblast tissue of the inner cell mass of an early stage embryo called a blastocyst. Studies conducted in the mid-1980s showed that these lines are valuable vehicles to enable targeted mutagenesis. Although, homologous recombination frequencies in mouse ES cells are higher in part because of their shortened cell cycle, gene targeting strategies remain low rate and low throughput. As of 2005, nearly 3600 mouse genes were mutated with this targeting strategy. This covers only less than15 % of the mouse genome **(94)**.

Although numerous factors influence homologous recombination frequency, the problems largely stem from the use of targeting vectors and ES cells with different genetic backgrounds. ES cells from the 129 mouse strain were the first to be derived and have become more widely used. Nevertheless, significant genetic variation exists among the different substrains. Before learning of this diversity, the first generation of gene targeting vectors were constructed from 129/SvJ DNA for use in 129/Sv ES cells. The sequence diversity between vector DNA from 129/SvJ and genomic DNA of 129/Sv ES cells may have reduced targeting efficiency and made targeting impossible. Some experiments targeting Tyr in Chr 7 reflected this problem **(41)**. Now, the only alternative to the 129 ES cell strain is C57/Bl6. However, its robustness for mutagenesis has not yet been demonstrated. This means that 129 ES cell are still the only tool for mutagenesis studies, and efforts should now be directed to extract more information from genomes of 129 substrains and expand low cost 129 substrain BAC libraries.

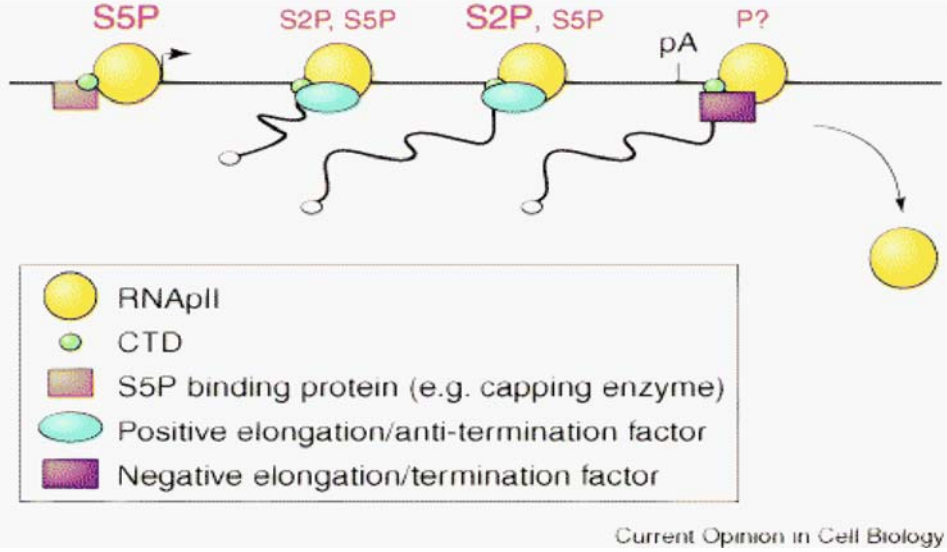
1.3 Transcription termination in mammals

Without any known exceptions, all living cells carry their hereditary information in DNA. DNA is transcribed into RNA, and this RNA code is translated to produce proteins, depending on a cell's need. In mammals, in addition to the tight control over transcription elongation, transcription termination also has complex regulation. Transcription termination is the process that first releases the transcription complex and then RNA polymerase from the template DNA. A critical reason for building the FLEx

cassette in this project is the need to reduce nonproductive transcription and prevent transcriptional interference at downstream genes, processes that are both linked to polyadenylation. It was found that both of these processes depend on the same DNA sequences placed 3' end of the eukaryotic protein coding genes. Some studies indicate that polyadenlylation factors and mRNA cleavage processes are involved in transcription termination **(121)**. Cleavage-polyadenylation signals are the cis elements required for normal termination in mammals. This was the good evidence of the relationship between two these processes.

Two general models, the anti-terminator model and the torpedo model, hypothesize how transcription termination and cleavage-polyadenylation affect each other. Although, they show mechanistical differences in lower and higher eukaryotes, the basics are the same in both organisms. Recently a hybrid model, the anti-terminator-torpedo model of RNAi , explained some of the paradoxes of anti-terminator and torpedo models **(6)**.

For the anti-terminator model, the presence of poly(A) signals at the mRNA 3^{I} end, 10-30nucleotides upstream of the cleavage site, initiates conformational changes in the factors associated with elongation **(Figure 21)**.



Buratowski 2005

Figure 21 Anti-terminator model

Transcription termination is caused by conformational changes of the RNA pol II complex. The presence of a polyadenylation signal leads to the dissociation of nascent RNA from the template DNA, either with the release of anti-terminator factors and/or termination factors. CTD phosphorylation shows a changeable pattern in every stage.

As for the torpedo model, which has the same logic of bacterial rho dependent transcription termination, cleavage of the transcript at the polyadenylation site generates a new 5^{I} end that is rapidly degraded by a 5^{I} -3^{I} exonuclease. This nuclease enzyme somehow enhances RNA polymerase II elongation complex destabilization results in termination **(Figure 22)**.

Compared to lower eukaryotes like yeast, some mammalian genes have two cleavage sites to ensure efficient transcription.

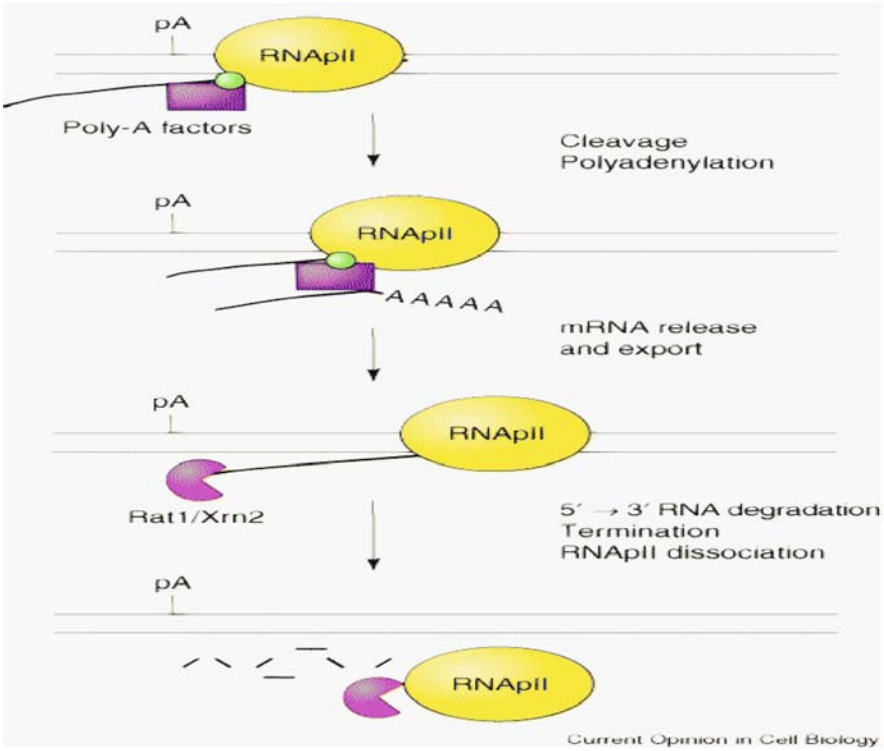


Figure 22 The torpedo model

Cleavage of the transcript by endonuclease at the polyadenylation region generates a 5' unprotected end. This is the target site for a 5^{I} -3^{I} exonuclease, Rat1 for yeast, Xrn2 for mammals. Via exonuclease activity, degradation of RNA downstream of the cleavage sites triggers RNA pol II dissociation, resulting in termination.

1.3.1 The presence of a CoTC element in some mammalian genes makes transcription termination through the torpedo model more powerful and efficient

RNA processing signals are critical for the transcription termination. Nevertheless, some studies have shown that there are more specific sites positioned downstream of the poly(A) that contribute to this process **(26,116)**.

In mammals, some genes, such as human β-globin, have a CoTC site several hundred nucleotides downstream of the polyadenylation signal. These sequences are required for auto-cleavage of nascent RNA transcript at this site. The 5^{I} end generated by the CoTC cleavage acts as a second entry point for the 5^{I} -3^{I} mammalian exonuclease

Xrn2. This exonuclease then passes along the RNA, reaches the 3^{I} RNA region and finishes the degradation. This action destabilizes the polymerase complex and dissociates it.

1.3.2 Pausing of the polymerase by G-rich sites, MAZ4

In addition to cleavage process contribution to transcription termination, some researchers have shown that pausing of the polymerase downstream of the polyadenylation site encourages transcription termination. Pausing can be triggered by particular DNA/RNA sequences or sequence specific DNA binding proteins that act as a roadblock to RNA polymerase II movement.

It has been demonstrated that pausing is critical for transcriptional termination of some yeast and mammalian genes. In *Schizosaccharomyces pombe,* pause sites adjacent to the pA in ura 4 and 200bp downstream of pA in nmt2 genes are required for transcription termination (*S.pombe* pause sites). The human α2 globin gene and C2 complement, as well as mouse immunoglobulin genes, display downstream elements that serve as pausing factors. Pausing downstream of the pA site is an integral part of the torpedo model of termination. Transcription complexes are slowed by pausing sites and exonuclease activity catches up them to release the template.

The presence of G rich specific sequences positioned downstream of the strong pA signal in human C2 complement was shown to be the binding site for the zinc finger protein, MAZ, a transcriptional regulation protein **(3)**. It was thought that the four MAZ protein binding sites arranged in tandem promoted termination of human C2 complement. Moreover, it activated polyadenylation. MAZ4 element mediation of transcription termination is dependent on proximity to a strong pA. RNAi-mediated depletion of MAZ protein indicated that there is no effect of the MAZ4 region on pausing activity. Therefore other factors or the specific G rich sequences, either at DNA or RNA level, may act to promote Pol II pausing. Although the mechanism of action of G rich pausing elements is not known yet, it might be explained by conformational change of a loop in the G rich sequences after the transcription complex and RNA polymerase II enzyme is

released from the template. On the other hand, the proximal distance of this region to the pA may act to retain RNA polymerase II in close proximity to the poly(A) site. When cleavage takes place, Pol II may then be close enough to be affected by Xrn2, 5^{I}-3^{I} exonuclease that is also involved in pause-dependent transcriptional termination **(36)**.

2. RESULTS

2.1 Generic FLEx gene trapping cassette and its operation principle

In the present study, we planned a multipurpose conditional-first, knockout strategy that is applicable to the majority of mouse genes regardless of their complex structure. Our starting consideration was that the majority of mammalian genes have relatively large intronic regions compared to the exons. The strategy was thus focused on inserting a cassette into an early, ideally first, intronic region of the target mouse gene in such as way that it did not interfere with downstream transcript production unless Cre recombinase was used to mediate a unidirectional cassette inversion. The construction of a versatile FLEx gene trapping cassette was the first aim of my PhD project. During the early phase of the cassette development study, the basic pre-mRNA processing elements, synt sA **(84)**, synt pA **(55)**, and transcription termination signals, CoTC **(26,116)** and/or MAZ4 **(3,36),** formed the core region of the cassette flanked by inversely oriented heterotypic lox site pairs, loxP **(53)** and lox5171 **(53)**. Then, the selectable marker was inserted between the synthetic sA and pA **(Figure 23)**.

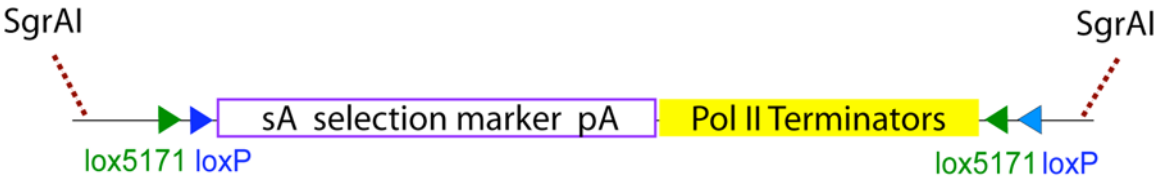


Figure 23 Architecture of generic FLEx gene trapping cassette for multi-purpose allele strategy

The FLEx cassette for conditional mutagenesis first needed to be neutral in one orientation, and following the Cre induction, mutagenic at the level of RNA processing by terminating transcription downstream of the cassette insertion. In this manner, this project is exploratory in using mammalian RNA polymerase terminators, CoTC and MAZ4, as neither of them has been applied in such a way before.

Considering the complexity of transcriptional control mechanisms in mammalian systems, understanding the individual functions of the cassette elements and their collaborative contributions to mutagenesis was the initial focus of my work. I wanted to explore two questions: What is the most effective distance between each element in the cassette? and, What effect does the character of the DNA between the splice acceptor and poly(A) signals have on efficient pre-mRNA processing events, splicing and polyadenylation? Therefore a series of FLEx cassettes containing elements in various configurations was engineered and tested in the first intronic region of pCAGGs-EGFP-SV40-Neo, a plasmid vector consisting of a strong promoter that is a hybrid of the CMV early immediate enhancer and a chicken β-actin promoter followed by a β-actin intron driving an enhancer green fluorescent marker fused to a neomycin selection marker under the control of the SV40 promoter. The cloning strategies used to assemble the first series of FLEx cassettes are detailed along with the results in figures 24, 25, 26 and 27.

Figure 24, section A, summarizes the PCR cloning strategy of the lox regions from 4.3 kb pBLF5171 into the NotI(670) and KpnI(772) cloning sites of 2.9 kb pBluescriptKS(+/-). The BsIWI restriction enzyme recognition site is CGTACG, and it generates a 5^I overhang (GTAC). The KpnI restriction enzyme recognizes GGTACC and it generates a 3^I overhang (GTAC). In this cloning strategy, the DNA fragments cleaved by these two enzymes were ligated. The colonies were selected by ampicillin (100 μg/ml). When the candidate recombinant plasmids were analysed with the SacI enzyme, recombinant plasmids, C4 andC5, were distinguished by their large size band compared to non-recombinant plasmids, C1-C3 and C6-C7 **(Figure 24B)**.

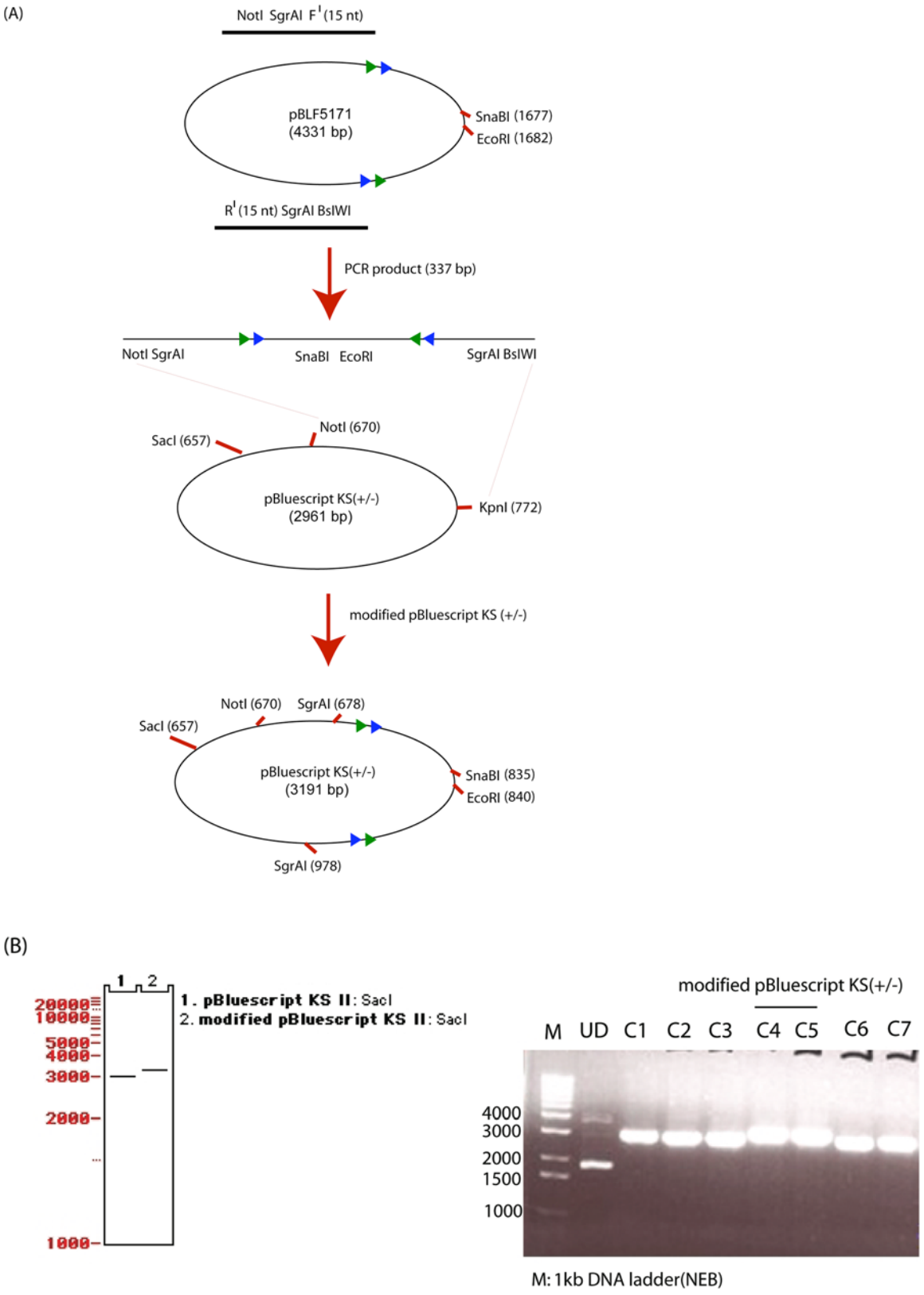


Figure 24 Subcloning of the lox regions to pBluescript KS(+/-)

After sequencing confirmed that the modified pBluescript KS(+/-) carried lox p and lox 5171 sites, synthetic sA, pA, and the transcription termination signal CoTC were inserted by PCR cloning between the SnaBI(835) and EcoRI(840) restriction sites of this plasmid **(Figure 25)**. For this study, the pUC18/human β-globin gene (3538 bp), kindly provided by Prof. Proudfoot, was the source of the CoTC fragment. The CoTC fragment of human β-globin lays between the genomic region coordinations 64567-64938 on chromosome 11. A 3.1 kb fragment of human β-globin gene, which spans the genomic region coordinates 62179-65316 on chromosome 11, was inserted into the HindIII cloning site of pUC18 (2686 bp) **(2,25)**.

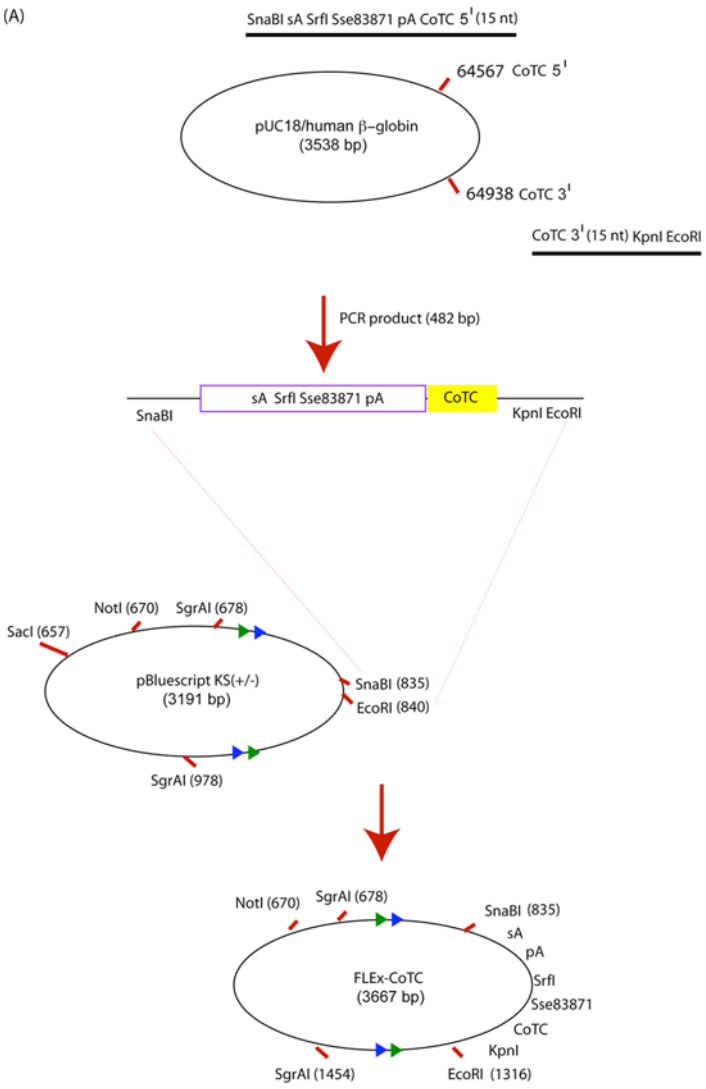


Figure 25 (A) Subcloning strategy of the pre-mRNA processing elements, synt sA, pA and transcription termination signal, CoTC.

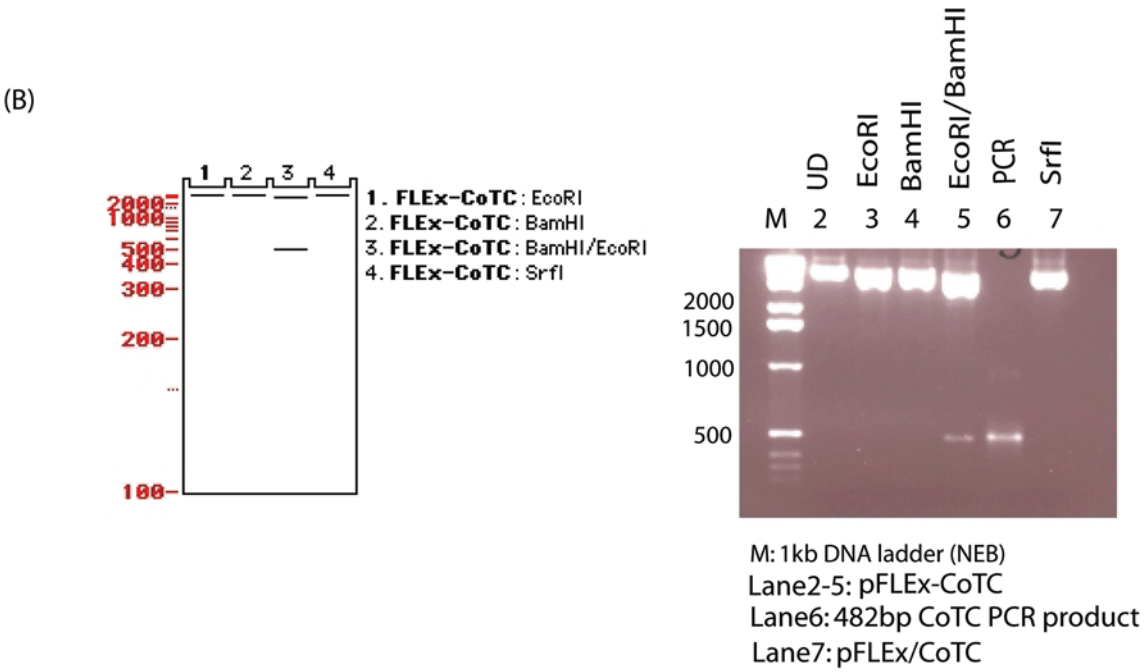


Figure 25 (B) Verification of pFLEx-CoTC

Following the modification of 3.191 kb pBluescript KS(+/-) through PCR cloning of sA, pA and CoTC, the 3.661kb recombinant plasmid was named pFLEx-COTC. pFLEx-CoTC was analysed by several restriction endonuclease enzymes (listed on gel). 482bp CoTC PCR product was used as a control.

The third subcloning step was to engineer the FLEx gene trapping cassette to contain two RNA pol II termination signals. The 150 bp MAZ4 RNA pol II pausing signal was inserted into the KpnI(1314) and EcoRI(1316) cloning sites of the pFLEx-CoTC **(Figure 26 A)**. pMLP III, kindly provided by Prof. Proudfoot, was used as a source of MAZ4 pausing region from the human complement genes C2 and factor B **(116)**. In this plasmid, the wild type 4 G nucleotides repeat sequences site, MAZ4, was placed into the region downstream of the synt pA **(Figure 26 A)**. pMLPIII was constructed by the insertion of the pUC19 polylinker to the pAdML (Stratagene). MAZ4 was inserted into the KpnI and EcoRI cloning sites of the pMLP III. Then, synt pA was cloned between the BamHI and KpnI sites of pMLP III **(116)**.

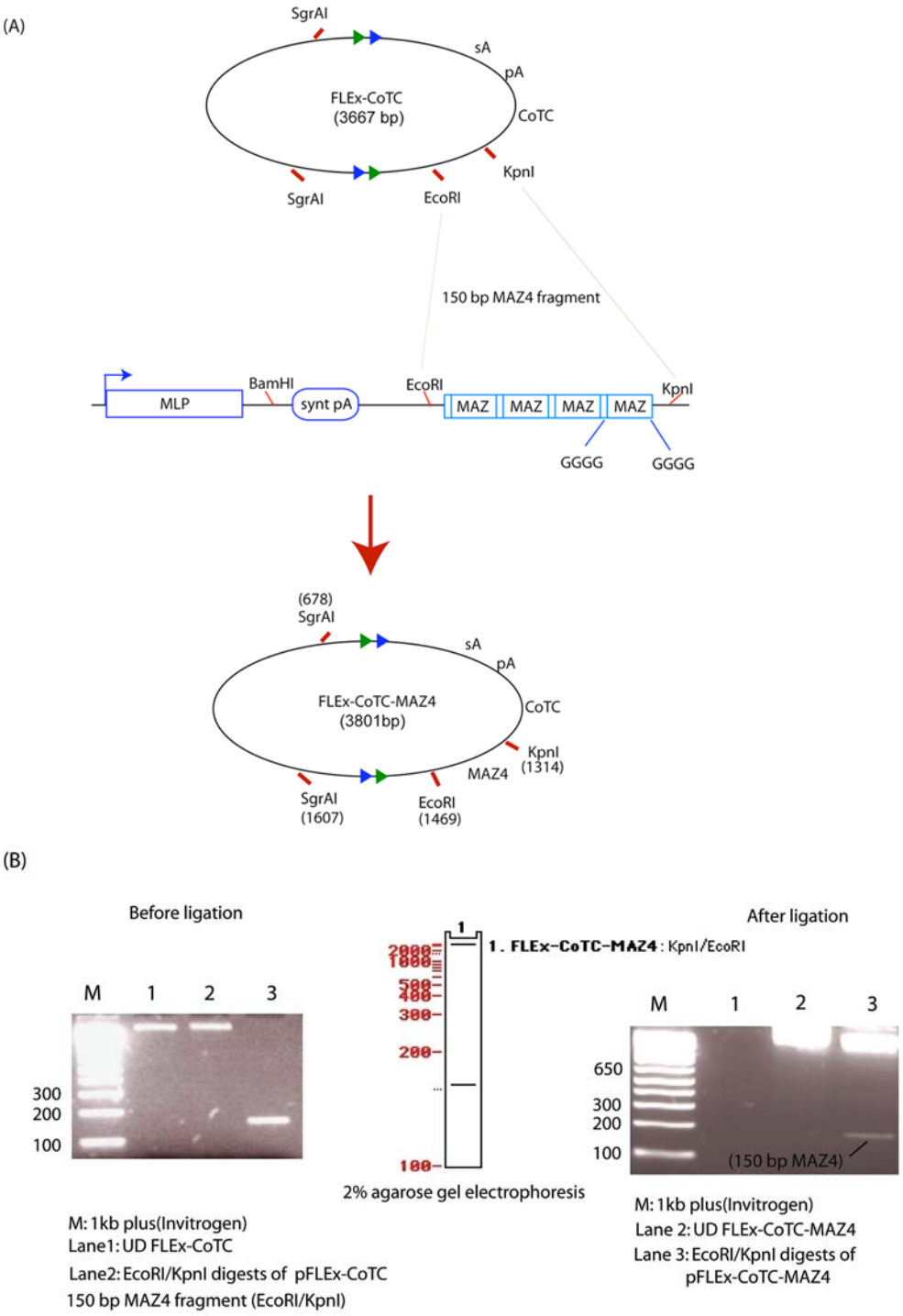


Figure 26 Modification of FLEx-CoTC construct via cloning of MAZ4
(A) Strategy of vector construction.
(B) 2% agarose gel electrophoresis of the KpnI/EcorI digestion of FLEx-CoTC-MAZ plasmid, 150kb. The sequencing facility verified correct modification of pFLEx-CoTC.

In order to determine the extent that the MAZ signal in the FLEx gene trapping cassette contributes to transcription termination, the CoTC site was removed from the 3.8 kb pFLEx-CoTC-MAZ4. First, the plasmid was digested with the double enzyme system, AatII and Asp718, an isoschizomer KpnI. Asp718 generated a blunt end product, so the AatII site was blunt ended by 3^{I} overhang removal and fill-in of 3^{I} recess using Klenow (DNA pol I), and the plasmid was then re-ligated **(Figure 27)**.

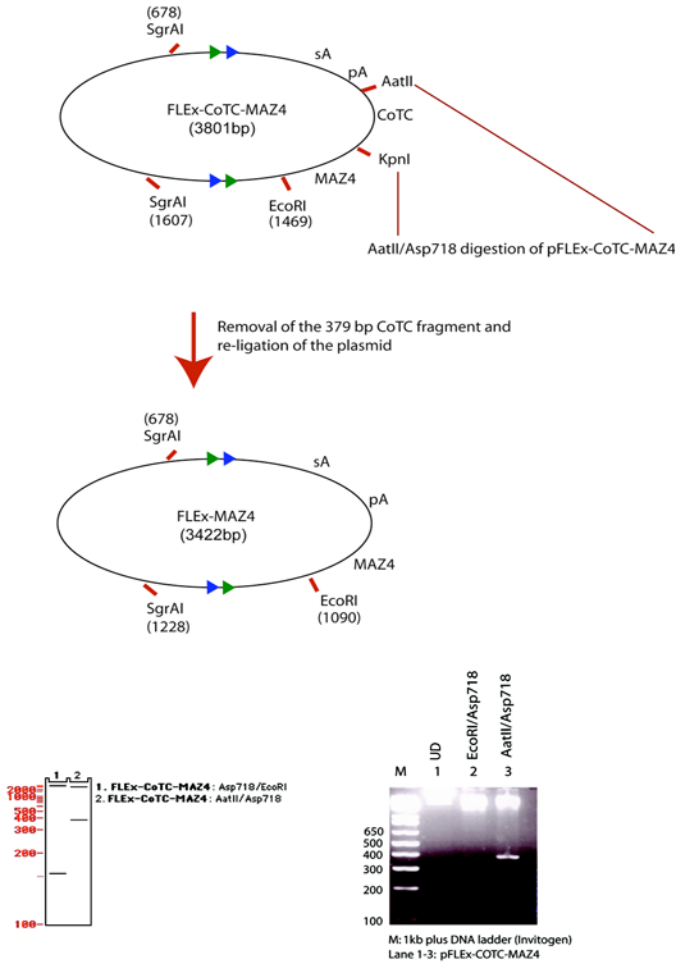


Figure 27 Construction of pFLEx-MAZ4

2.1.1 Evidence for communication between pre-mRNA processing factors, splicing, and polyadenylation

The first series of FLEx cassettes constructed with synthetic sA, synthetic pA and different configurations of the transcription termination signals CoTC and MAZ4, was cloned into the β-actin intron region of the pCAGGs-EGFP-SV40-Neo in a bidirectional fashion. This was done to generate neutrality in one direction and a knock-out, at the level of GFP transcriptional control, in the other **(Figure 28)**.

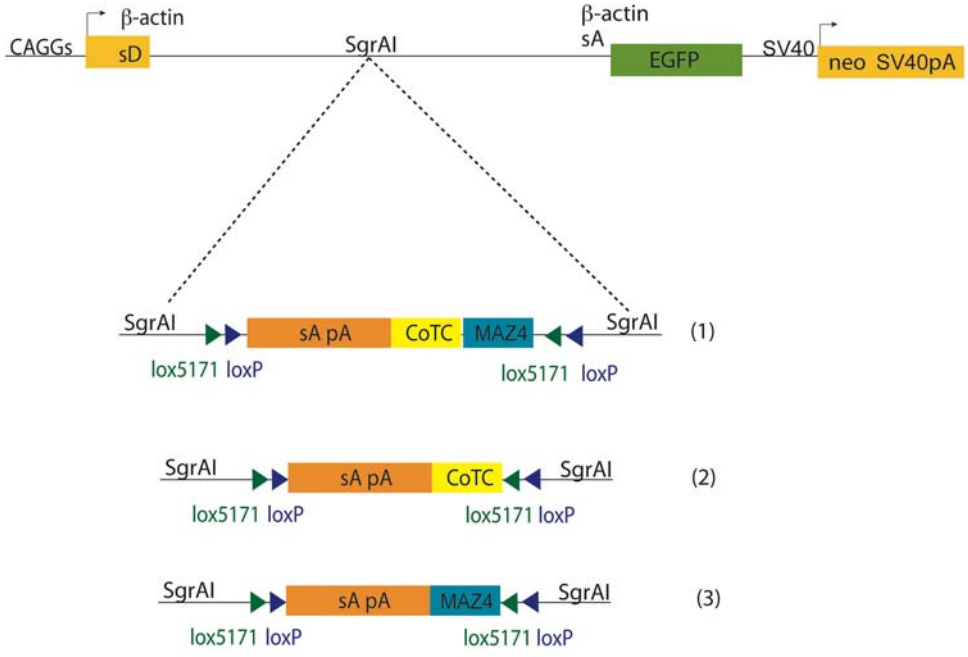


Figure 28 Bi-directional cloning of the FLEx cassette and its two different sets

Cloning strategy for the FLEx cassette composed of synthetic sA, pA, and transcription termination signals CoTC and MAZ4 to the unique SgrA1 site in the β-actin region of pCAGGs-EGFP-neo. (2) and (3) illustrate different configurations of the FLEx cassette in terms of termination signal usage. In one set, only CoTC was used as a terminator, in the other, MAZ4 was used. FLEx cassettes and its two different series were inserted into the intronic region in both directions; CAGGs promoter direction and CAGGs anti-promoter direction.

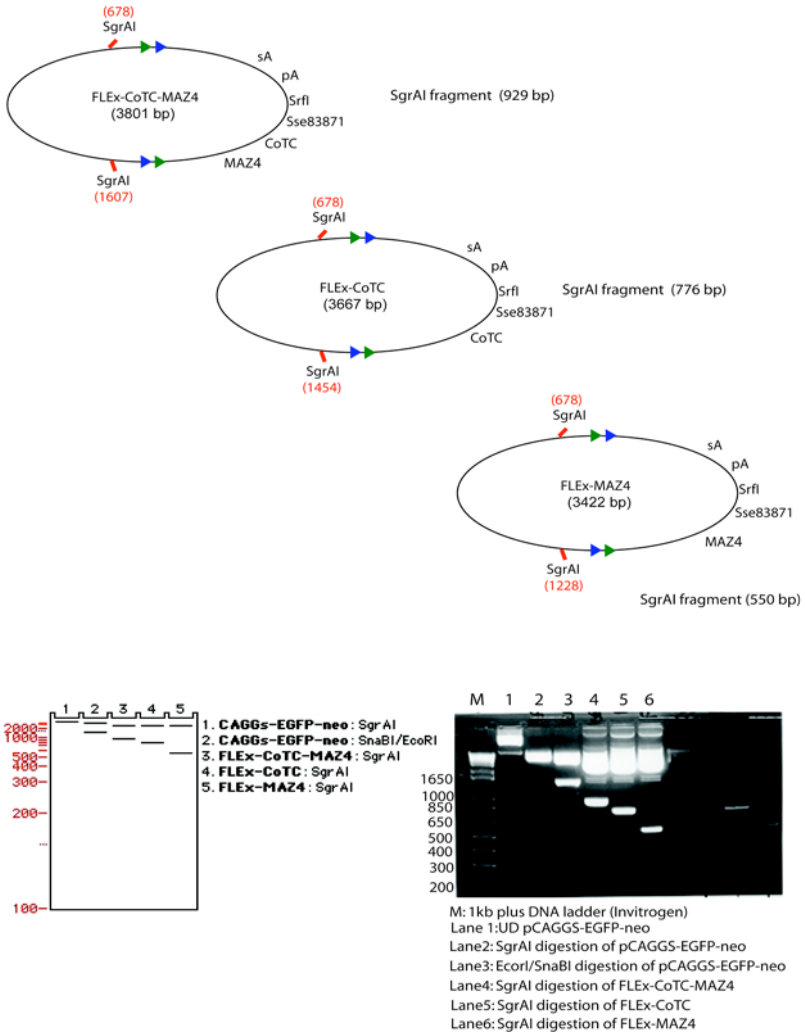


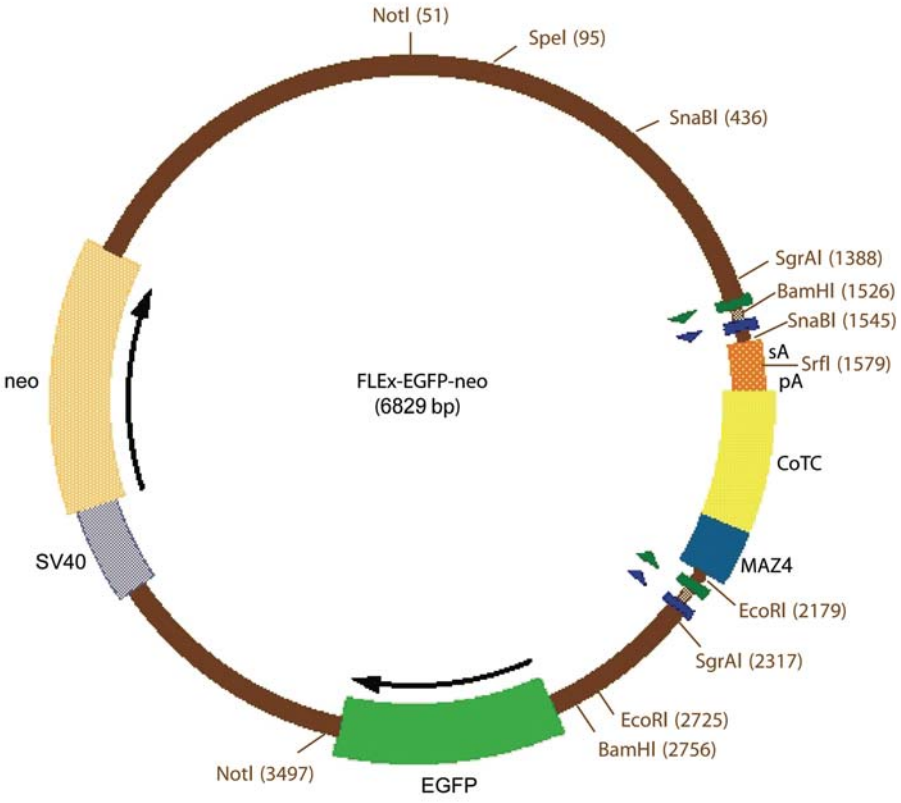
Figure 29 Insertion of the FLEx cassette into the β-actin intronic region of pCAGGs-EGFP-neo

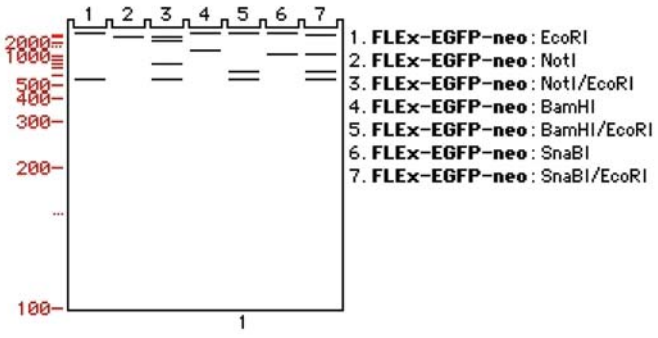
SgrAI fragments of pFLEx-CoTC-MAZ4 and its two variations were cloned into the SgrAI cloning site of the β-actin intronic region of pCAGGs-EGFP-neo (upper figure). 0.7% agarose electrophoresis of SgrAI fragments of the FLEx constructs and pCAGGs-EGFP-neo is shown (lower figure).

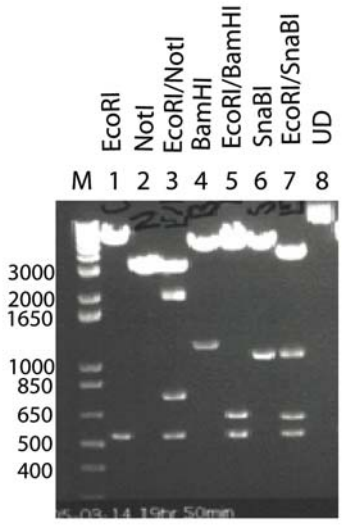
6 different FLEx-EGFP-neo plasmids were generated with this cloning strategy. These included FLEx cassettes composed of the two RNA pol II termination signals (CoTC and MAZ4) and two variations constructed with either MAZ4 or CoTC inserted into the intronic region of pCAGGs-EGFP-neo in two directions, the CAGGs promoter direction and the anti-CAGGs promoter direction. Plasmids with the CAGGs promoter orientated FLEx gene trapping cassettse were called “mutagenic oriented pFLEx-EGFP-

neo". Plasmids with anti-CAGGs promoter oriented FLEx cassettes were called "neutral oriented rev-FLEx-EGFP-neo". In the presence of CoTC and MAZ4, 6.8 kb pFLEx-EGFP-neo and rev-FLEx-EGFP-neo constructs were built and verified by restriction enzyme analysis and sequencing.

The restriction patterns of the mutagenic oriented FLEx-EGFP-neo construct with CoTC and MAZ4 are: EcoRI, **546** and **6283 bp**; NotI, **3383** and **3446 bp**; EcoRI and NotI double digestion, **546, 772, 2128 and 3383 bp**; BamHI, **1230 and 5599 bp**; EcoRI and BamHI double digestion, **653 and 6176 bp**; SnaBI, **1109 and 5720 bp**; and, EcoRI and SnabI double digestion, **546, 634, 1109 and 4540 bp (Figure 30)**. The restriction patterns of the neutral oriented FLEx-EGFP-neo construct with CoTC and MAZ4 are: BamHI, **546 and 6283 bp**; SgrAI, **929 and 5900 bp**; EcoRI and SnaBI double digestion, **561, 638, 1090 and 4540 bp**; SrfI and SpeI linearized **6829 bp**; and, EcoRI and NotI, **772,1199,1475 and 3383 bp (Figure 31)**.







M: 1kb plus DNA ladder (Invitrogen)
Lane1-8: mutagenic oriented FLEx-EGFP-neo

Figure 30 Map and agarose gel electrophoresis of several digests of pFLEx-EGFP-neo before ES cell transfection

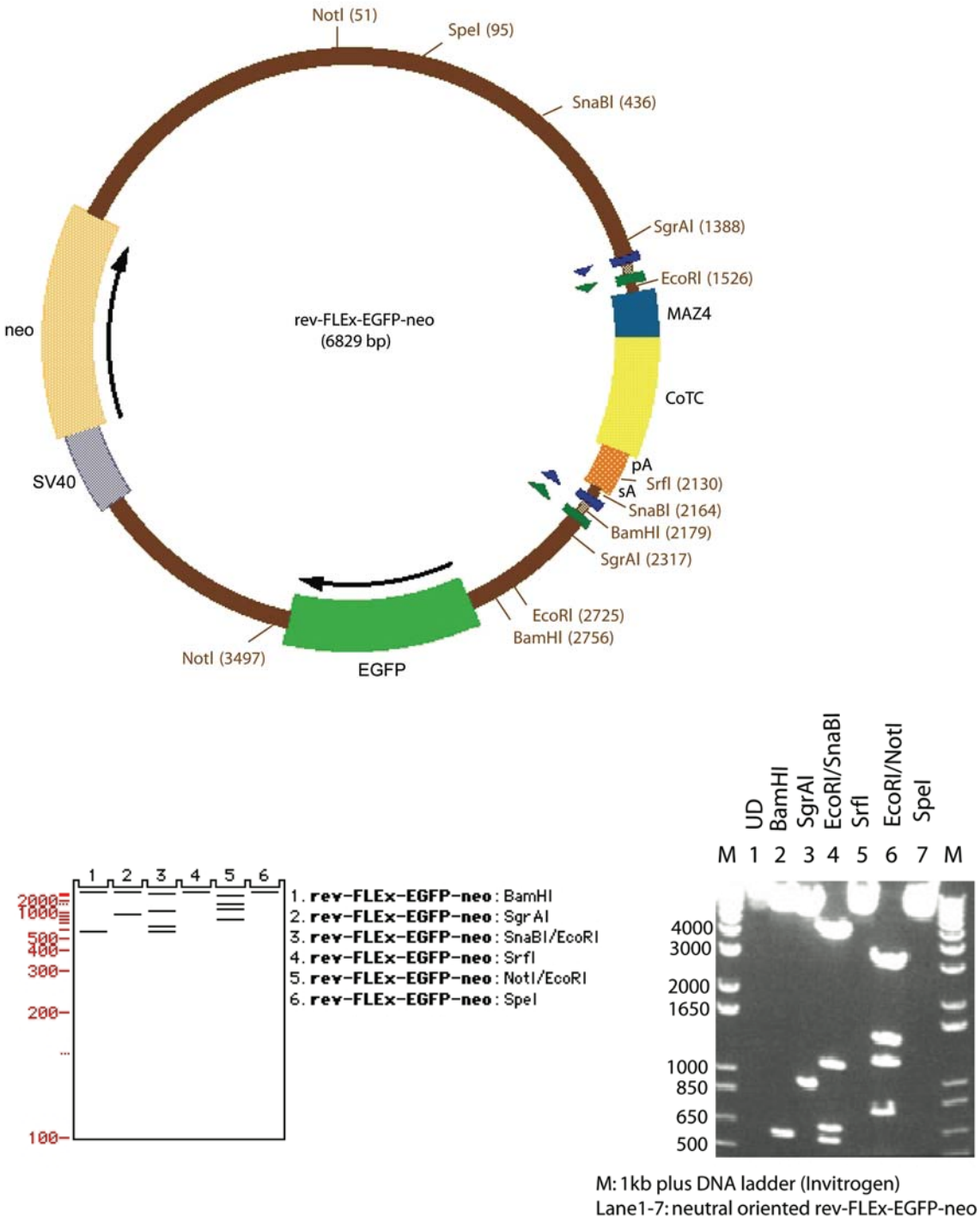


Figure 31 Map and agarose gel electrophoresis of several digests of rev-FLEx-EGFP-neo before ES cell transfection

The restriction patterns of the mutagenic oriented FLEx-EGFP-neo construct with a single CoTC are: BamHI, **1077 and 5599 bp**; SgrAI, **776 and 5900 bp**; EcoRI and SnaBI double digest, **481, 546, 1109 and 4540 bp**; EcoRI and NotI double digestion, **546, 772, 1975 and 3383 bp**; and, SrfI and SpeI both linearized, **6676 bp** plasmid **(Figure 32, left gel)**. The restriction patterns of the neutral oriented FLEx-EGFP-neo construct with CoTC are: BamHI, **577 and 6099 bp**; SgrAI, **776 and 5900 bp**; EcoRI and SnaBI double digestion, **485, 561, 1090, and 4540 bp**; EcoRI and NotI double digestion, **772, 1046, 1475, and 3383 bp**; and, SrfI and SpeI both linearized, **6676 bp (Figure 32, right gel)**.

The restriction patterns of the mutagenic oriented FLEx-EGFP-neo construct with MAZ4 are: BamHI, **855 and 5599 bp**; SgrAI, **557 and 5897 bp**; EcoRI and SnaBI double digestion, **259, 546, 1109 and 4540 bp**; and, EcoRI and NotI double digestion, **546, 772, 1753 and 3383 bp (Figure 33, left gel)**. The restriction patterns for the neutral orientation of this construct are: BamHI, **577 and 5877 bp**; SgrAI **557 and 5897 bp**; EcoRI and SnaBI double digestion, **263, 761, 1090 and 4540 bp**; and, EcoRI and NotI double digestion, **772, 824, 1475 and 3383 bp**. SrfI and SpeI linearized both constructs to gave a **6454 bp** fragment.

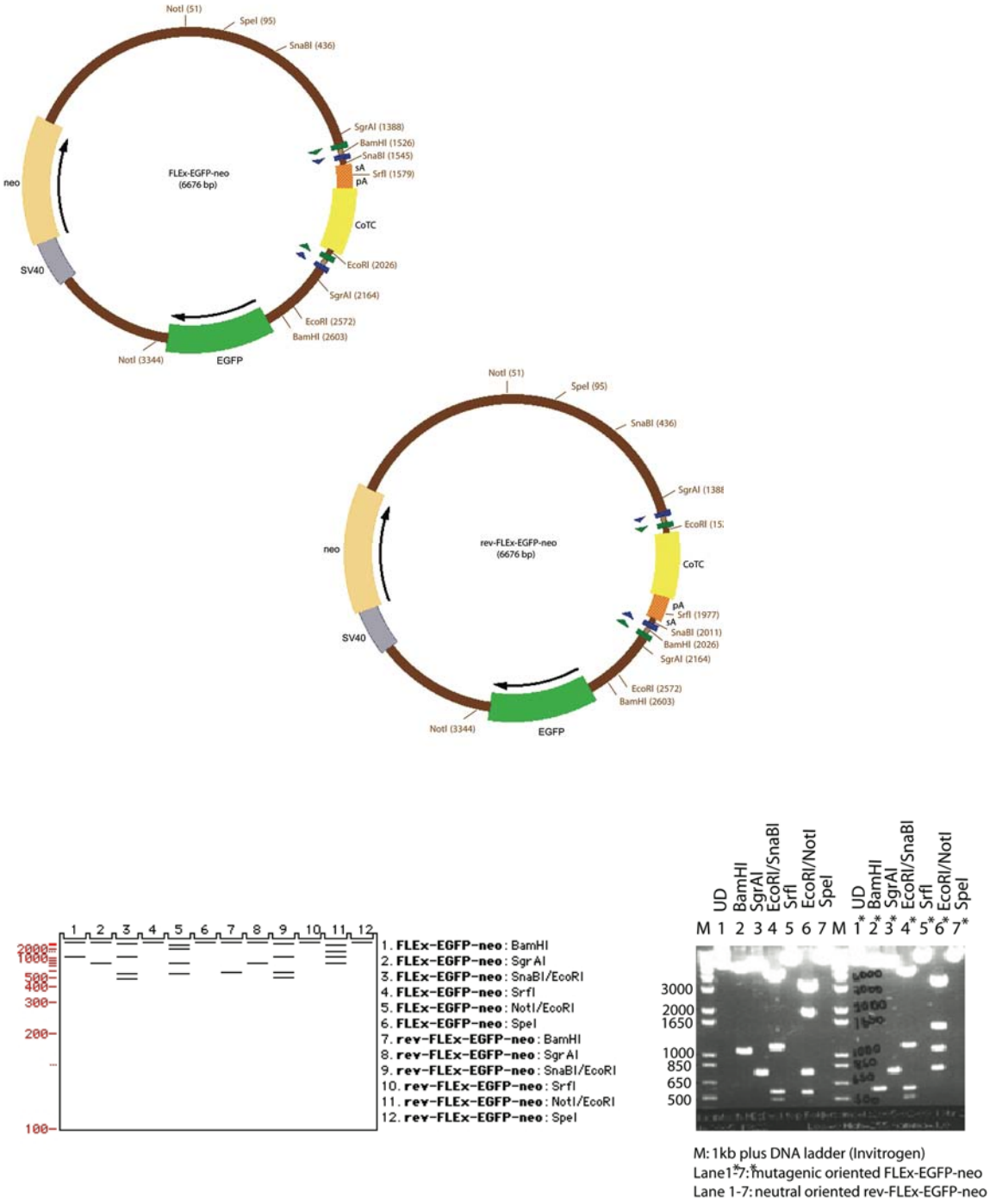


Figure 32 Verification of mutagenic and neutral oriented FLEx-EGFP-neo constructs that include the RNA pol II termination signal CoTC by analysis with several restriction endonucleases.

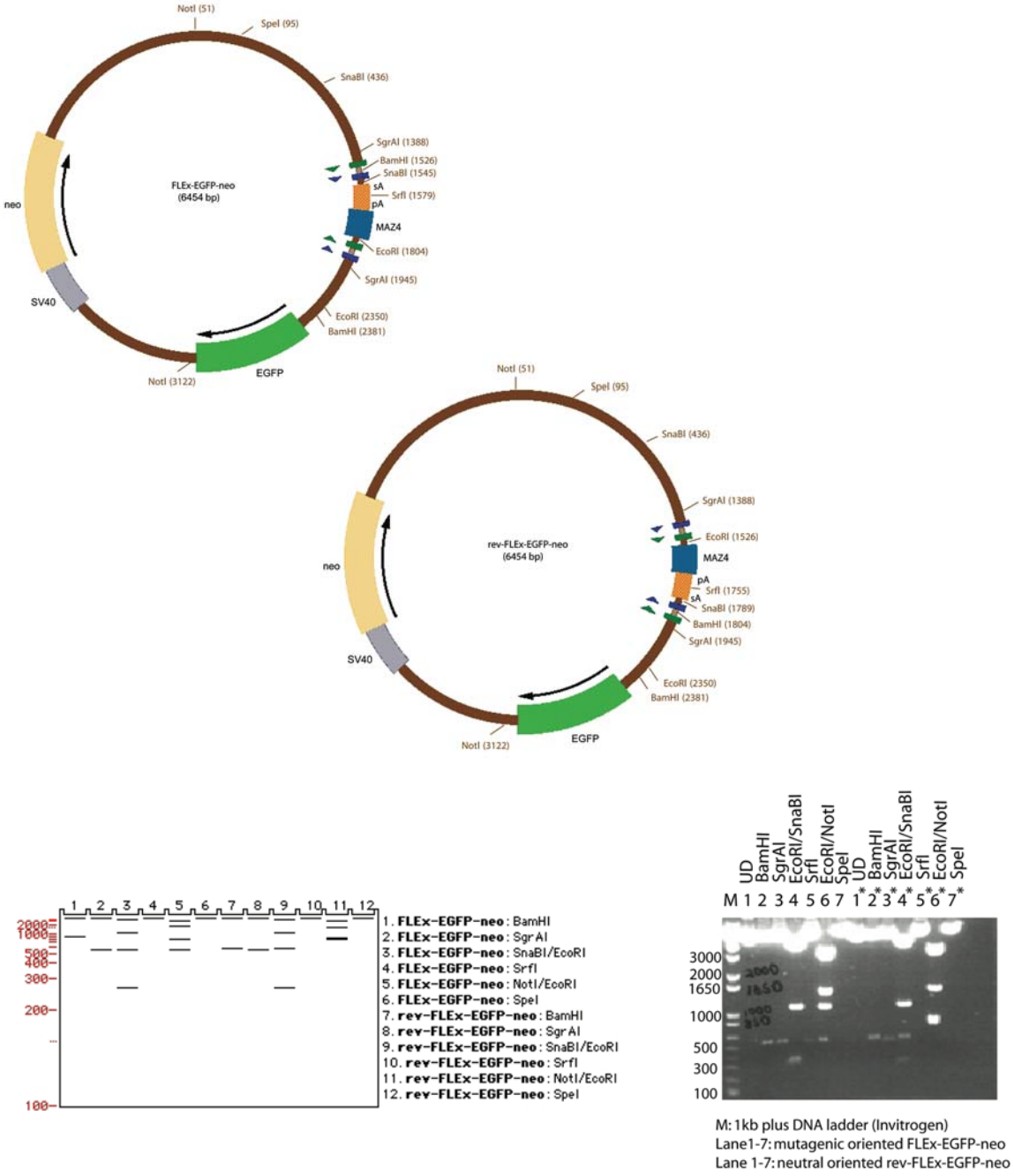


Figure 33 Verification of mutagenic and neutral oriented FLEx-EGFP-neo constructs containing the RNA pol II pausing signal MAZ4, using restriction endonuclease analysis.

Random mutant mouse E14TG2a cells were established by electroporating the FLEx-EGFP-Neo vector series, linearized with SpeI, on the backbone of the pCAGGs-EGFP-neo. Two weeks following electroporation, and continuous selection of the clones

with G418 (0.20mg/ml), the ES clones were examined using a fluorescent microscope. Statistical data were obtained from 40 recombinant clones for each construct. The fluorescent signals of the cells indicated whether the FLEx cassette randomly integrated into the mouse genome in a mutagenic or neutral position. As shown by the positive GFP expression the cassette in the neutral orientation was reliable in all vector series **(Figure 34)**.

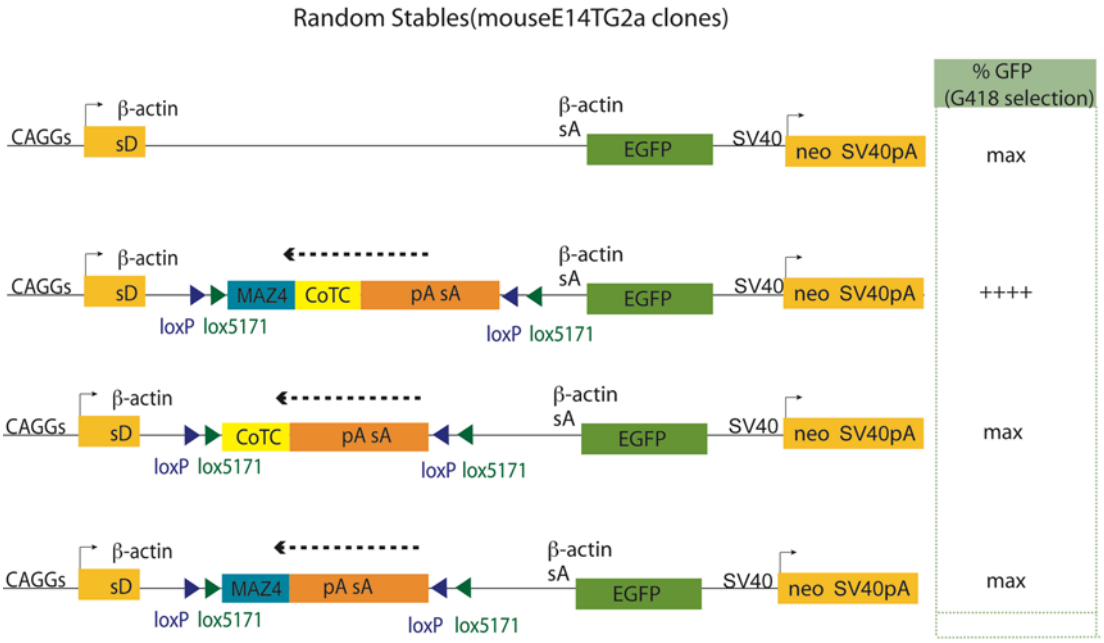


Figure 34 GFP assay results of the neutral oriented first series of FLEx-EGFP-neo mouse E14TG2a clones

Fluorescent microscopy analysis of the random stable E14TG2a clones. 3x10 cm plates/ 40 clones for each transfection were subjected to this assay. Clones expressing GFP were classified as 0-15% GFP(+), 15-30%(++), 30-45%(+++), 45-60%(++++) and >60% (max). Comparison to the control group (max GFP expression) indicated that all cassette configurations did not interfer with downstream GFP transcription.

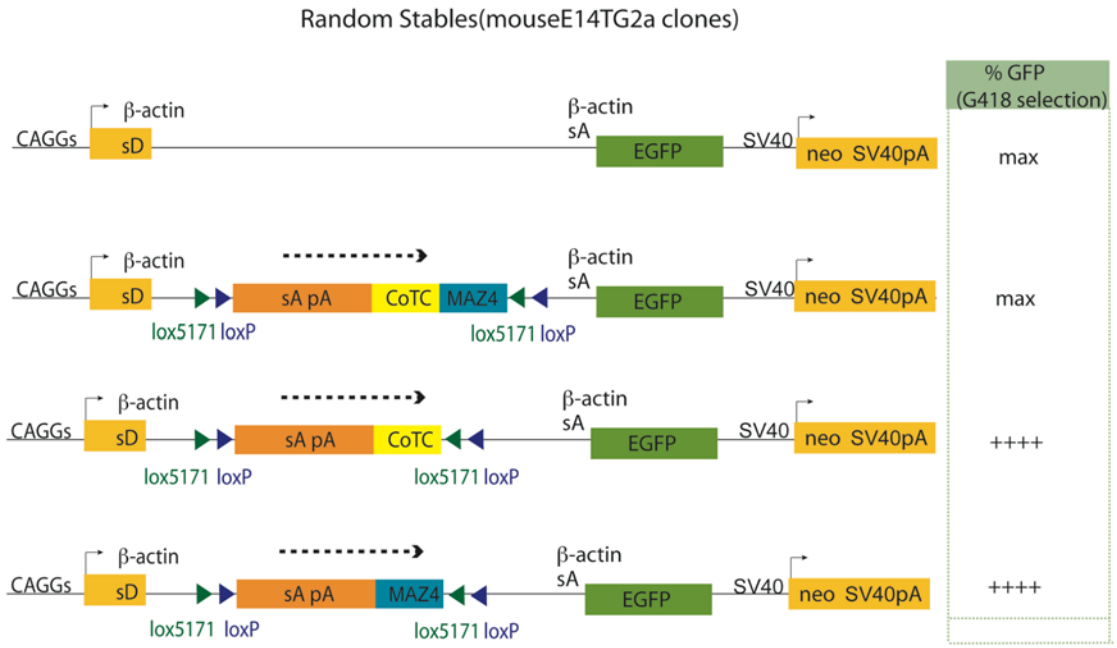


Figure 35 (A) GFP assay results of mutagenic oriented first series of FLEx-EGFP-neo mouse E14TG2a clones

Fluorescent microscopy analysis of the random stable E14TG2a clones. 3x10 cm plates/ 40 clones for each transfection were subjected to this assay. Clones expressing GFP were classified as 0-15% GFP(+), 15-30%(++), 30-45%(+++), 45-60%(++++) and >60% (max). Compared to the control group (max GFP expression), the GFP assay results held by using the clones transfected with mutagenic oriented 3 sets of FLEx cassette indicated the cassette in every configuration could not generate mutation at the level of RNA processing, GFP was expressed in every condition.

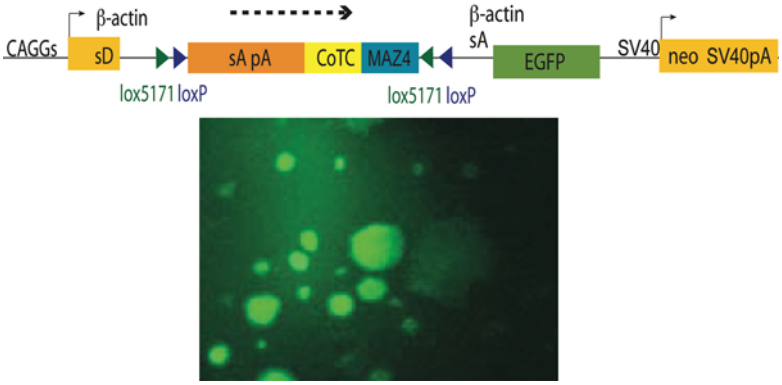


Figure 35 (B) *In vivo* transcriptional activity of mutagenic oriented FLEx –EGFP-neo containing both termination signals, CoTC and MAZ4, in mouse E14TG2a clones/10cm plate

On the other hand, in CAGGs promoter direction, it did not show fidelity for mutation leading to a transcription stop at the downstream region of the FLEx cassette, EGFP gene **(Figure 35)**. This could be explained by random integration of the FLEx cassette into an active chromatin region of the mouse genome or by inefficient splicing. To explore this problem further, some recombinant clones and control clones were picked from 96 well plates and expanded into 6 well plates. A RT-PCR assay was performed on these clones to determine if the cassette was skipped by the splicing machinery. The results showed splicing failure of the cassette in some clones **(Figure 36).**

Total RNA concentration from random stables(E14Tg2aclones)				
Clones	GFP expression	Sample absorbance(260nm)	C	Reverse transcribed total RNA
CAGGS-EGFP-neo (control)	(+)	0.144	0.720μg/μl	1.7μl(1.2μg)
CAGGS-EGFP-neo (control)	(-)	0.069	0.345μg/μl	3.6μl(1.2μg)
mutagenic FLEx-EGFP-neo(both signals)	(+)	0.059	0.295μg/μl	4.2μl(1.2μg)
mutagenic FLEx-EGFP-neo(both signals)	(-)	0.184	0.920μg/μl	1.4μl(1.2μg)
neutral FLEx-EGFP-neo (both signals)	(+)	0.088	0.440μg/μl	2.8μl(1.2μg)
neutral FLEx-EGFP-neo (both signals)	(-)	0.081	0.405μg/μl	2.9μl(1.2μg)
mutagenic FLEx-EGFP-neo(CoTC)	(+)	0.223	1.115μg/μl	1.1μl(1.2μg)
mutagenic FLEx-EGFP-neo(CoTC)	(-)	0.172	0.860μg/μl	1.4μl(1.2μg)
neutral FLEx-EGFP-neo (CoTC)	(+)	0.182	0.910μg/μl	1.3μl(1.2μg)
neutral FLEx-EGFP-neo (CoTC)	(-)	0.102	0.510μg/μl	2.5μl(1.2μg)
mutagenic FLEx-EGFP-neo(MAZ4)	(+)	0.116	0.580μg/μl	2.2μl(1.2μg)
mutagenic FLEx-EGFP-neo(MAZ4)	(-)	0.068	0.340μg/μl	3.5μl(1.2μg)
neutral FLEx-EGFP-neo (MAZ4)	(+)	0.081	0.405μg/μl	2.9μl(1.2μg)
neutral FLEx-EGFP-neo (MAZ4)	(-)	0.153	0.765μg/μl	1.3μl(1.2μg)

Table 4 Total RNA concentrations used for RT-PCR assay

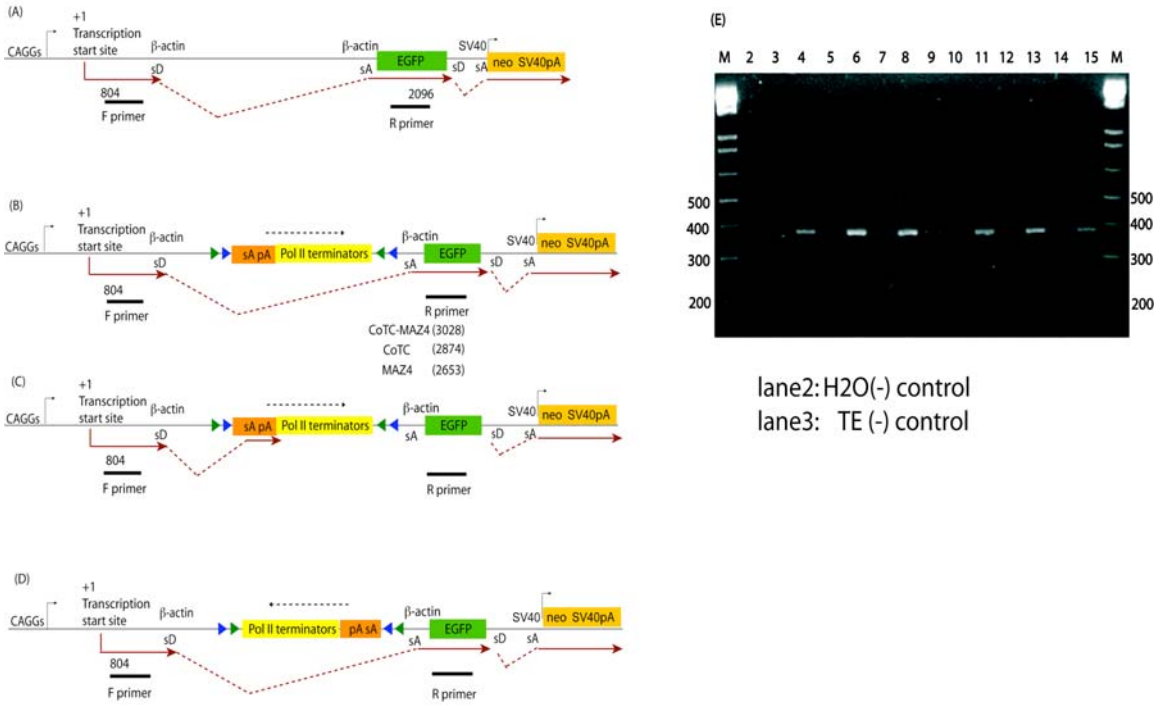


Figure 36 RT-PCR analysis of GFP expressions in random stables (recombinant mouse E14TG2a clones)

(A) In control clones transfected with pCAGGS-EGFP-neo, following the transcription, (transcripts were shown as red arrows), the 912 bp β- actin intronic region between the β-actin sD to sA spliced from 1292 bp transcript within the specific RT-PCR primers boundaries ; forward primer from the 100 bp downstream of the transcription start site and reverse primer from the EGFP region. Thereby, these primers amplified the 380 bp EGFP cDNA product **(figE, lane 4)**. If the GFP was not transcribed, the cDNA could not be amplified **(figE,lane 5).**

(B) In mutagenic oriented FLEx cassette transfected clones, if the splicing machinery skipped the sA in the cassette, and β-actin region spliced from the β-actin sD to β-actin sA. In clones transfected with mutagenic oriented FLEx cassette with both signals, from 2224 bp transcript within the RT-PCR primers region, 1844bp was removed **(figE, lane 6)**. In clones stable with FLEx –CoTC-EGFP-neo constructs, from the 2067 bp transcript, 1687 bp was removed **(figE, lane 9)**. on the other hand for clones stable with FLEx-MAZ4-EGFP-neo, 1846 bp pre-mRNA produced within the primer borders and 1466 bp region spliced **(figE, lane 12)**.For all stable clones, 380 bp cDNA was amplified by RT-PCR primers.

(C) If the gene is spliced from the endogeneous donor to cassette splicing acceptor, the downstream EGFP can not transcribed and no product amplification was seen. RT-PCR result for random stable with mutagenic oriented FLEx with both signals **(figE lane 7)** FLEx with CoTC **(figE lane 10)** and FLEx with MAZ4(**fig E lane 13).**

(D) For the neutral oriented FLEx cassette transfected clones, the synthetic sA was not counted and the transcript spliced from the enodgeneous donor site to the endogeneous sA site.. 380 bp cDNA product was amplified by RT-PCR primers **(figE, lane 8 RT-PCR result from random stable with mutagenic oriented FLEx with both signals, lane 11 FLEx with CoTC and lane 14 FLEx with MAZ4)**

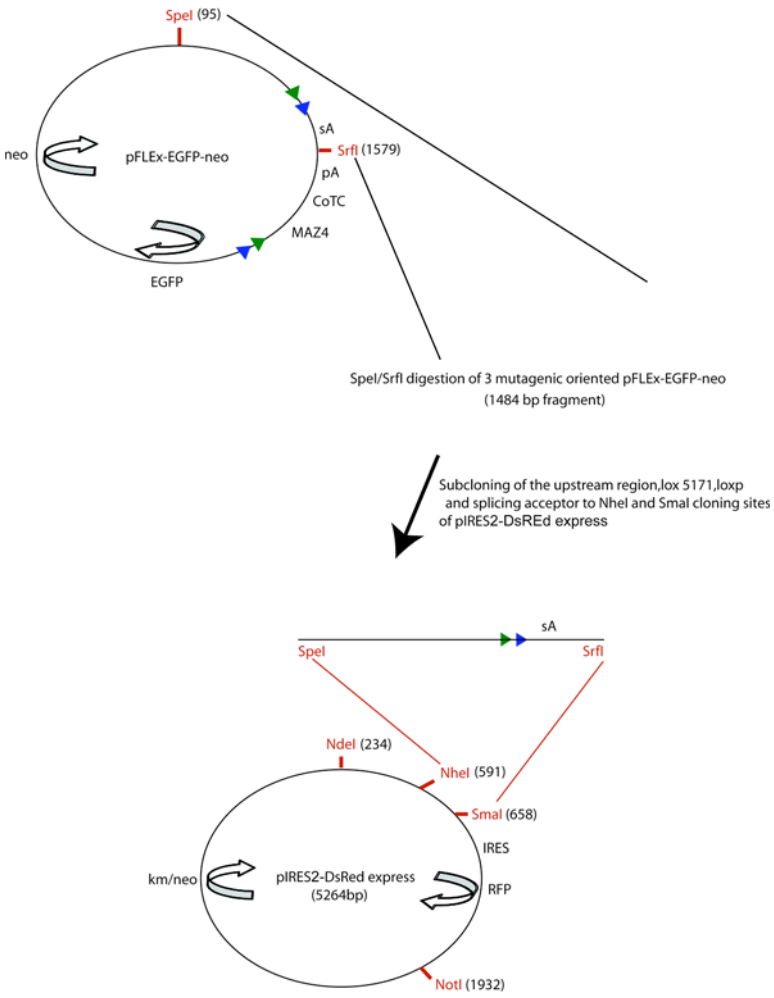


Figure 37(I) First step in the cloning strategy to build FLEx-IRES-RFP-EGFp-neo constructs

First, the upstream SpeI(95) and SrfI(1579) region, covering lox5171, loxP and the synthetic sA of the mutagenic oriented FLEx-EGFP-neo construct, was cloned into pIRES2-DsRed NheI(591) and SmaI(658) cloning sites.

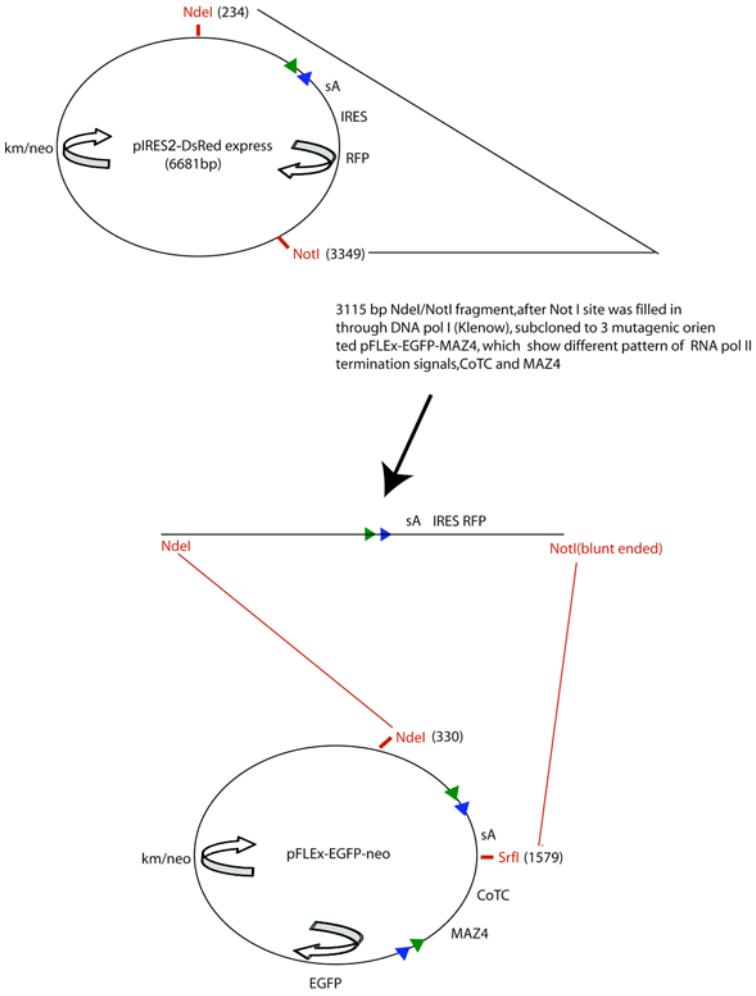


Figure 37(II) Second step in the cloning strategy to build FLEx-IRES-RFP-EGFP-neo constructs

Second, the aim was to add IRES-RFP between the synthetic spA and synthetic polyadenylation site of the FLEx-EGFP-neo constructs. The 3.1 kb Nde1-filled in NotI fragment of the modified pIRES2-DsREd construct was re-cloned to all three mutagenic pFLEx-EGFP-neo constructs (containing both CoTC and MAZ4, or only CoTC or MAZ4).

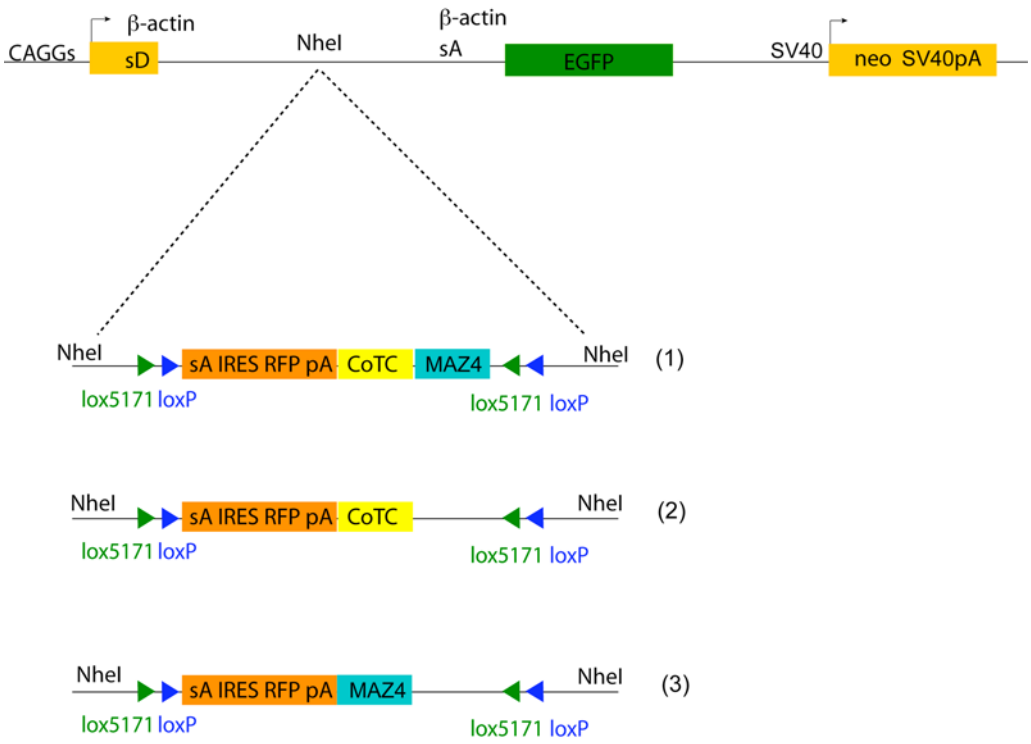


Figure 38 Neutral oriented FLEx-IRES-RFP-EGFP constructs were engineered through a bi-directional cloning approach into pCAGGs-EGFP-neo. The SgrAI cloning site of pCAGGs-EGFP-neo was replaced with NheI via Red-ET recombination. All mutagenic oriented FLEx-IRES-RFP-EGFP-neo SgrAI sites flanked by mutant and wild-type lox regions were changed through Red-ET recombination approach using NheI cloning sites.

We conclude that the splicing and polyadenylation machinery affected each other because of the distance between synthetic sA and synthetic pA, and this was the main reason for aberrant function of FLEx gene trapping cassette. Because of this, our next step was to design a second series of FLEx-EGFP-neo constructs called FLEx-IRES-RFP-EGFP-neo **(Figure 37)**. This second generation of FLEx cassettes included an in-frame IRES-RFP sequence to maintain an appropriate distance between the sA and the pA. This was confirmed by different restriction enzyme digests and sequencing **(Figure 39-41)**.

The restriction patterns of FLEx-IRES-RFP-EGFP-neo with both transcription termination signals are: EcoRI, **546 and 8153 bp**; BamHI, **55, 2453 and 6191 bp**; and, EcoRI and BamHI double digestion, **31, 55, 546, 1876 and 6191 bp**. For the neutral orientation of the same construct, the restriction patterns are: EcoRI, **2477 and 6222 bp**; BamHI, **55, 577 and 8067 bp**; and, EcoRI and BamHI double digestion, **31, 55, 546, 1876 and 6191bp (Figure 39)**.

The restriction patterns of FLEx-IRES-RFP-EGFP-neo containing CoTC are: EcorI, **546 and 7999bp**; BamHI, **55, 2299 and 6191 bp**; and, EcoRI and BamHI double digestion, **31, 55, 546,1722, and 6191 bp**. For the neutral orientation of the same construct, rev-FLEx-IRES-EGFP-neo, the restriction patterns are: EcoRI, **2323 and 6222 bp**; BamHI, **55, 577 and 7692 bp**; and, EcoRI and BamHI double digestion, **31, 55, 546, 1722 and 6191 bp (Figure 40)**.

The restriction patterns of FLEx-IRES-RFP-EGFP-neo containing MAZ4 are: EcorI, **546 and 7778 bp**; BamHI, **55, 2078 and 6191 bp**; and, EcoRI and BamHI double digestion, **31,55, 546, 1501 and 6191 bp**. The restriction patterns for the neutral oriented FLEx-IRES-RFP-EGFP containing MAZ4 are: EcoRI, **2102 and 6222bp**; BamHI, **55, 577, 7692 bp**; and, EcoRI and BamHI double digestion, **31, 55, 546, 1501, and 6191 bp (Figure 41)**.

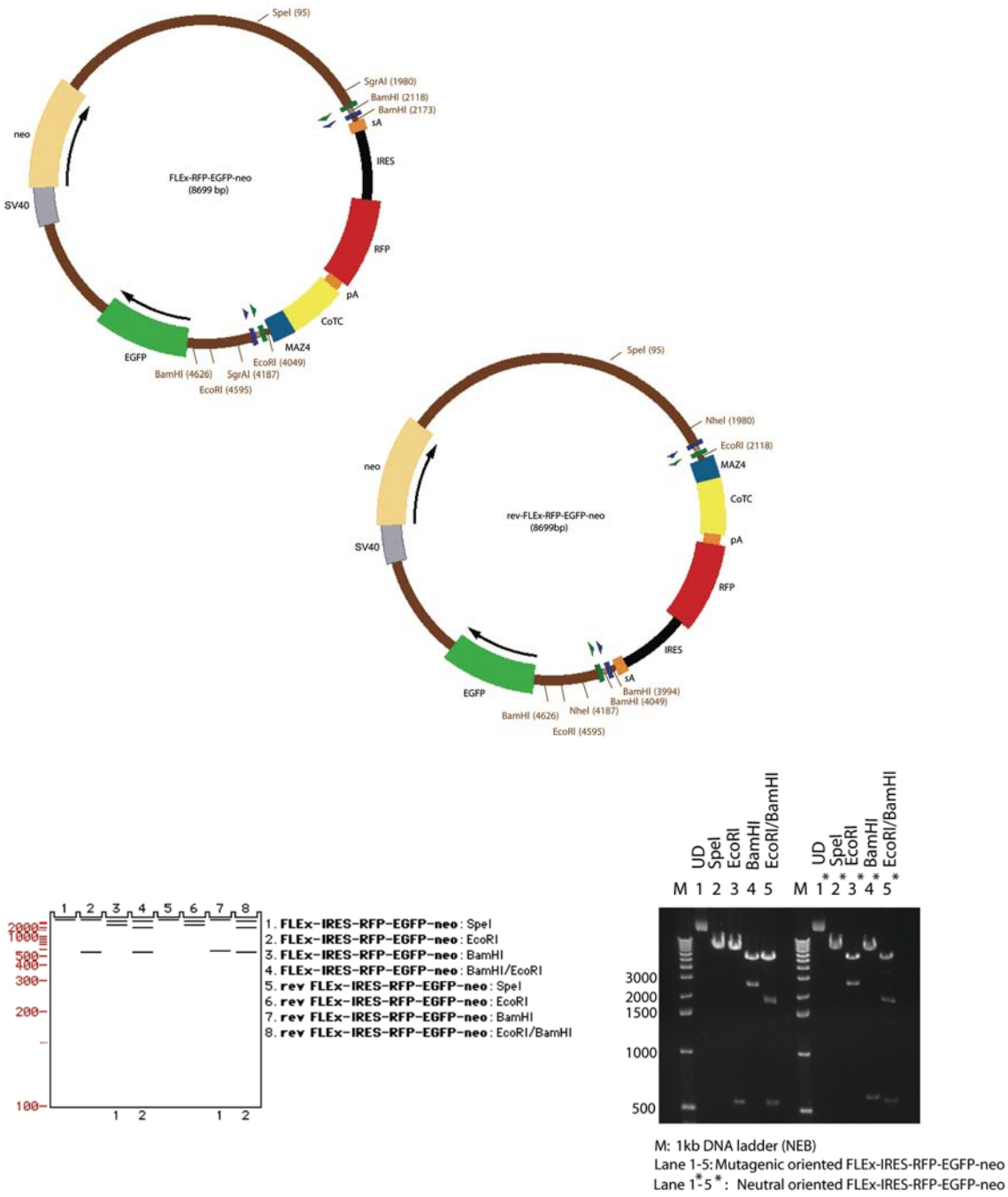


Figure 39 Map and agarose gel electrophoresis of several enzyme analytical digests of pFLEx-IRES-RFP-EGFP-neo in both orientations containing two RNA pol II signals, CoTC and MAZ4, before ES cell transfection

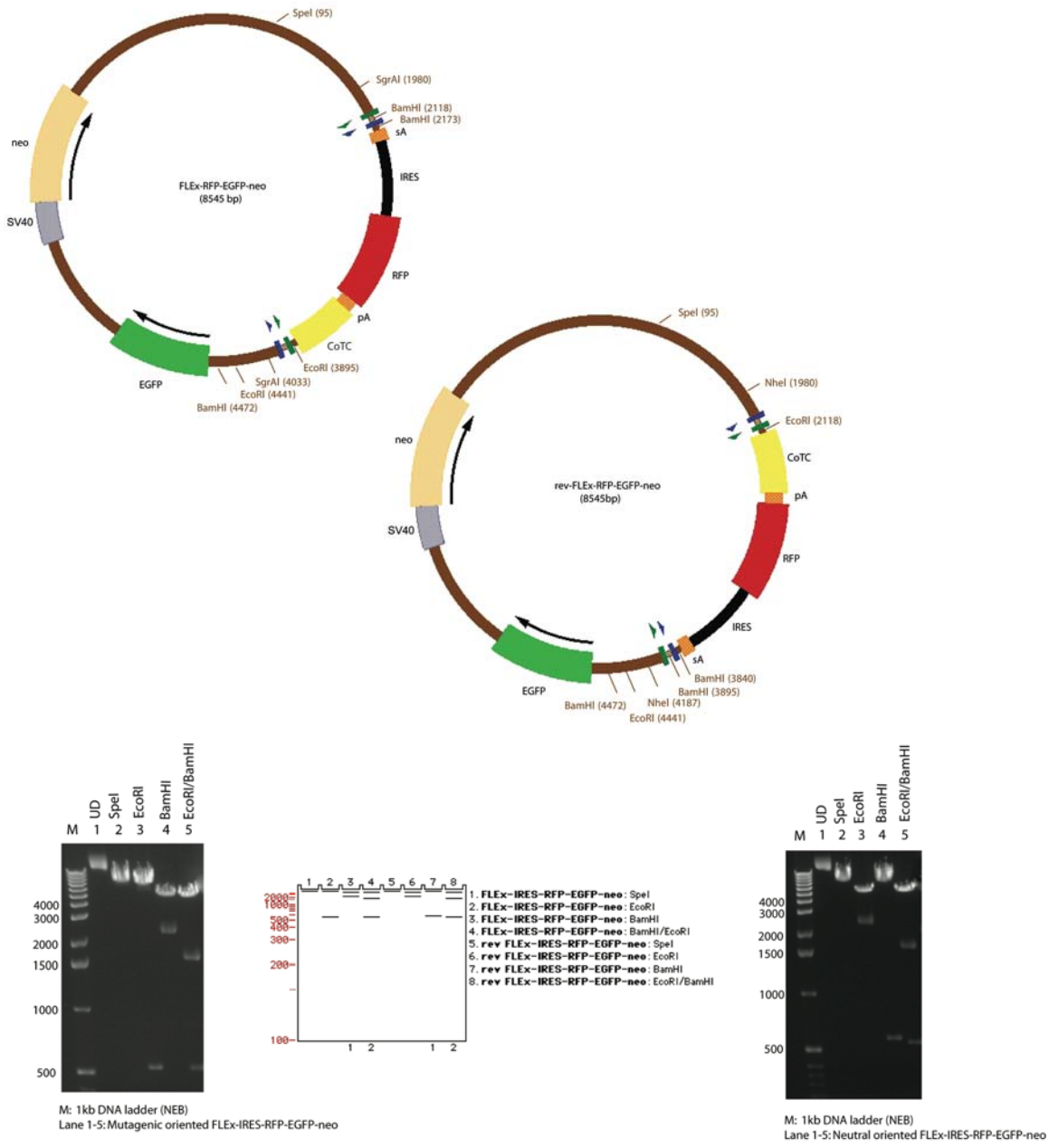


Figure 40 Map and agarose gel electrophoresis of several enzyme analytical digests of pFLEx-IRES-RFP-EGFP-neo in both orientations containing a single RNA pol II signal, CoTC, before ES cell transfection

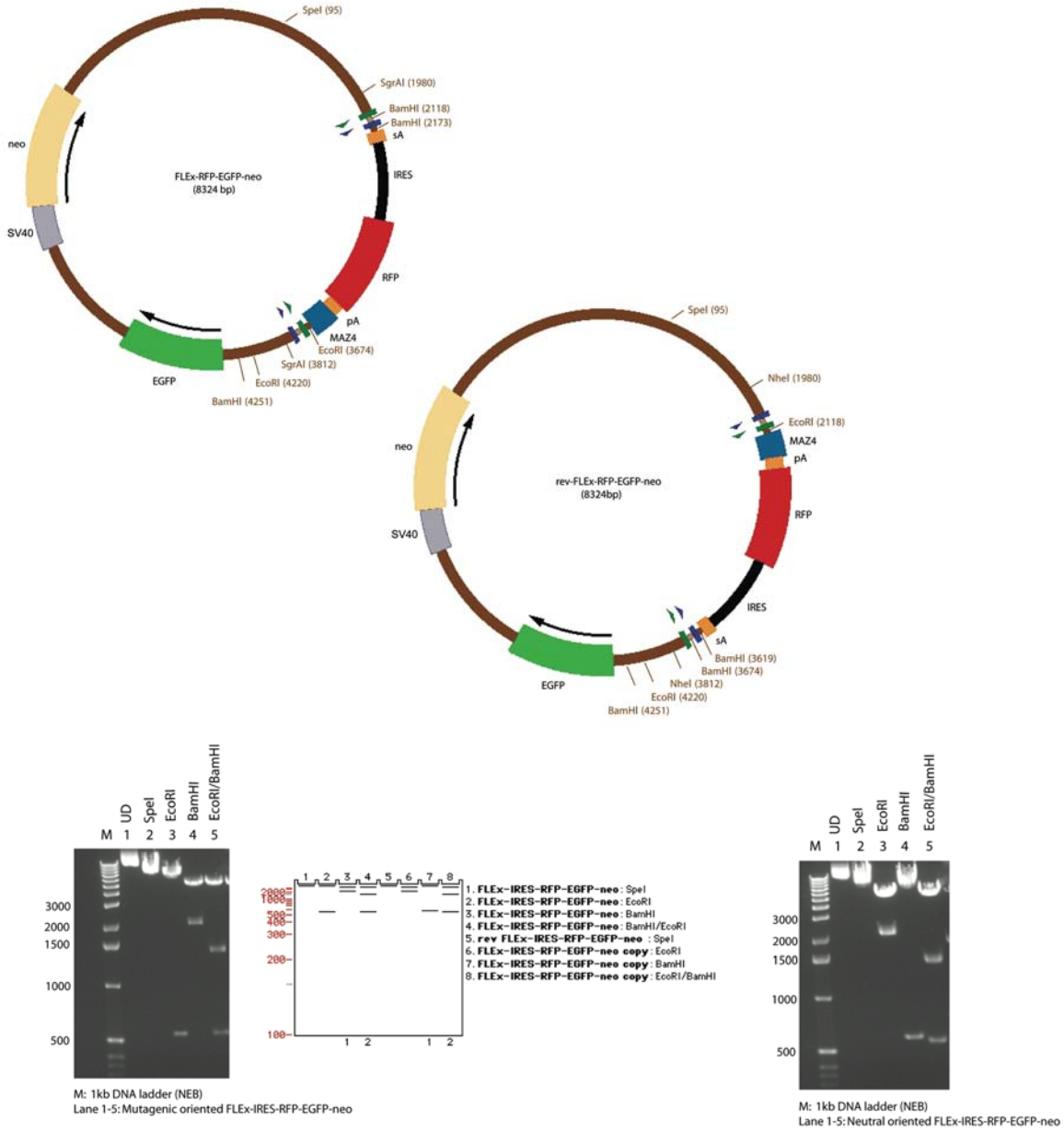


Figure 41 Map and agarose gel electrophoresis of several analytical digests of pFLEx-IRES-RFP-EGFP-neo in both orientations containing a single RNA pol II signal, MAZ4, before ES cell transfection

All constructs were linearized with SpeI, and E14TG2a cell lines were generated. After 14 days, G418 (0.20mg/ml) selection was performed and clones were examined under the fluorescent microscope. 40 clones were used for statistical analysis. The GFP assay results supported the distance requirement between the sA and poly(A). When the

linearized mutagenic oriented FLEx-EGFP-Neo vector integrated into the mouse genome, the synt sA and synt poly(A) elements served to respectively trap and truncate the transcript. This resulted in EGFP not being transcribed downstream of the cassette **(Figure 42)**.

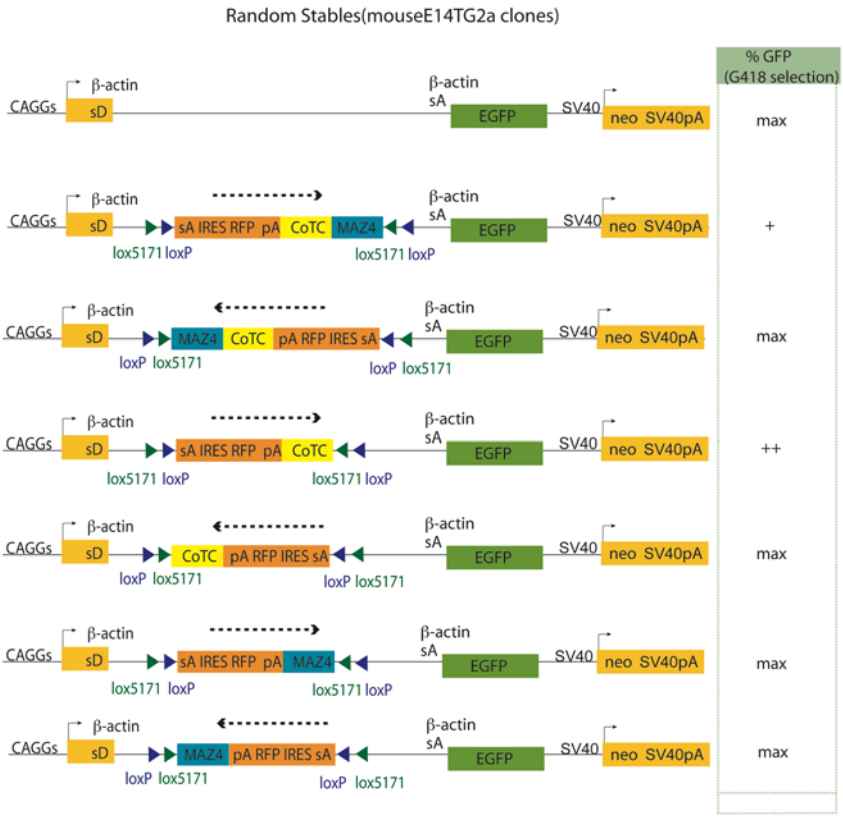


Figure 42 GFP assay results of mutagenic and neutral oriented second series of FLEx-EGFP-neo mouse E14TG2a clones

Fluorescent microscopy results for the random stable E14TG2a clones. 3x10 cm plates/ 40 clones for each transfection were subjected to this assay. Clones expressing GFP were classified as 0-15% GFP(+), 15-30%(++), 30-45%(+++), 45-60%(++++) and >60% (max). In the presence of both transcription termination signals, compared to MAZ4 alone (max, 65% of total clones), the CoTC containing cassette was more precise at controlling downstream transcription. (+) is equal to 12,5%,and (++) is equal to 20 % of total clones expressing GFP.

The second generation of the FLEx cassettes was then first tested by transient transfection of mouse E14TG2a ES cells. Lipofectamine was used for the vector transfection. Compared to stably transfected cells generated by electroporation, the mutagenic oriented FLEx cassette was more robust in these transiently transfected cells. We observed no trace of the GFP expression in these cells, however RFP was strongly expressed as expected. This confirmed the transient transfection of these cells. **(Figure 43)**.

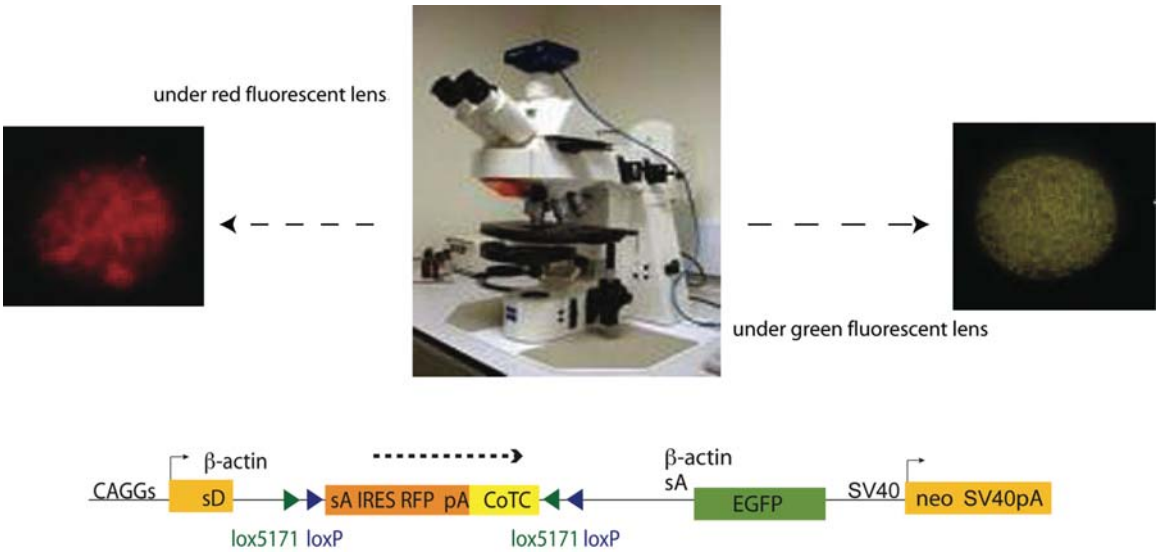


Figure 43 Transient transfection of mutagenic oriented FLEx-EGFP-neo construct

Examination of clones in a 6 well plate 24 hours after transient transfection of the vector illustrated (scale bar 40X). Under a lens that transmits red fluorescent signals, RFP was visualized (left side). However, under a green fluorescent filter, GFP signals were not detected (right side).

We also performed a Cre recombination feasibility test on selected clones stable for the 3 different versions of the mutagenic oriented FLEx cassette (FLEx-IRES-RFP-EGFP-neo with both (CoTC+MAZ4) or single transcription termination signal. This was done by transiently transfecting the pCAGGs-Cre plasmid, again using the lipofectamine approach. We found that the FLEx cassette was robust in terms of Cre mediated recombination **(Figure 44)**.

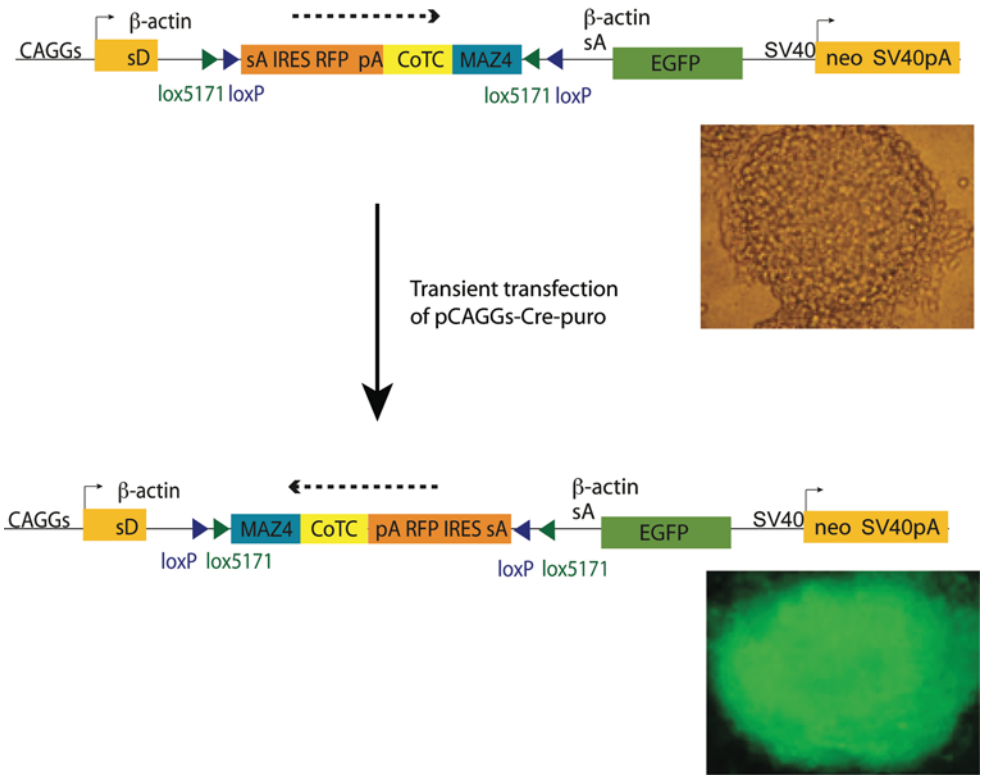


Figure 44 Transient transfection of pCAGGs-Cre-puro into the FLEx-CoTC-EGFP-neo clone

Cre efficiency test in a E14TG2a cell clone stable for the mutagenic oriented FLEx-EGFP-neo construct (scale bar 40X). Control clones were transiently transfected with pBluscript KS plasmid. After 24 hours the clone transiently transfected with Cre plasmid was positive for GFP expression, while the control was negative.

2.1.1.1 Conclusion

For this kind of applied gene trapping strategy, versatility of the gene trapping cassette is very important. With FLEx, our goal was to generate mutations at the RNA transcript processing level after inserting the cassette in a sense direction into an early intronic region of a target gene. Importantly, we found that the structure of the elements comprising the cassette, in addition to their position and proximity to each other, played a pivotal role in this mutagenesis. Our efforts during the first and second generation of the cassette were directed towards explaining splicing and polyadenylation interference when the sites are close together. We were particularly focused on the role of transcription

termination signals in preventing downstream gene transcription by RNA pol II. We derived two basic conclusions from our experimental efforts; sA and pA elements must be separated by a certain distance, and CoTC was more effective then MAZ4 in causing transcription termination.

2.1.2 Polyadenylation is required for transcriptional termination but is not sufficient

In order to determine the value of the transcription termination signals used in this study, the signals were removed from the two differentially oriented FLEx cassettes **(Figure 45)** and random stables were generated by electroporating these two vector constructs into mouse E14TG2a cells. Both FLEx-IRES-RFP-EGFP constructs containing only sA and pA elements were digested with SpeI, EcoRI and BamHI. The restriction patterns of the mutagenic oriented FLEx-IRES-RFP-EGFP construct were: EcoRI, **546 and 7629 bp**; and, BamHI, **55, 1929 and 6191 bp**. The **2453 bp** BamHI restriction fragment of mutagenic oriented FLEx-IRES-RFP-EGFP containing both CoTC and MAZ4 were used as a control. The restriction patterns of the neutral oriented FLEx-IRES-RFP-EGFP-neo construct not containing CoTC and MAZ4 were: EcoRI, **1873 and 6142 bp**; and, BamHI, **55, 497 and 7463 bp**. The **2477 bp** EcoRI fragment of the neutral oriented FLEx-IRES-RFP-EGFP-neo construct containing both CoTC and MAZ4 was used as a control **(Figure 45)**.

After 14 days, G418 selection (0.20mg/ml) was performed and positive GFP expression showed that these signals are required to create a mutation resulting in transcription termination **(Figure 46)**.

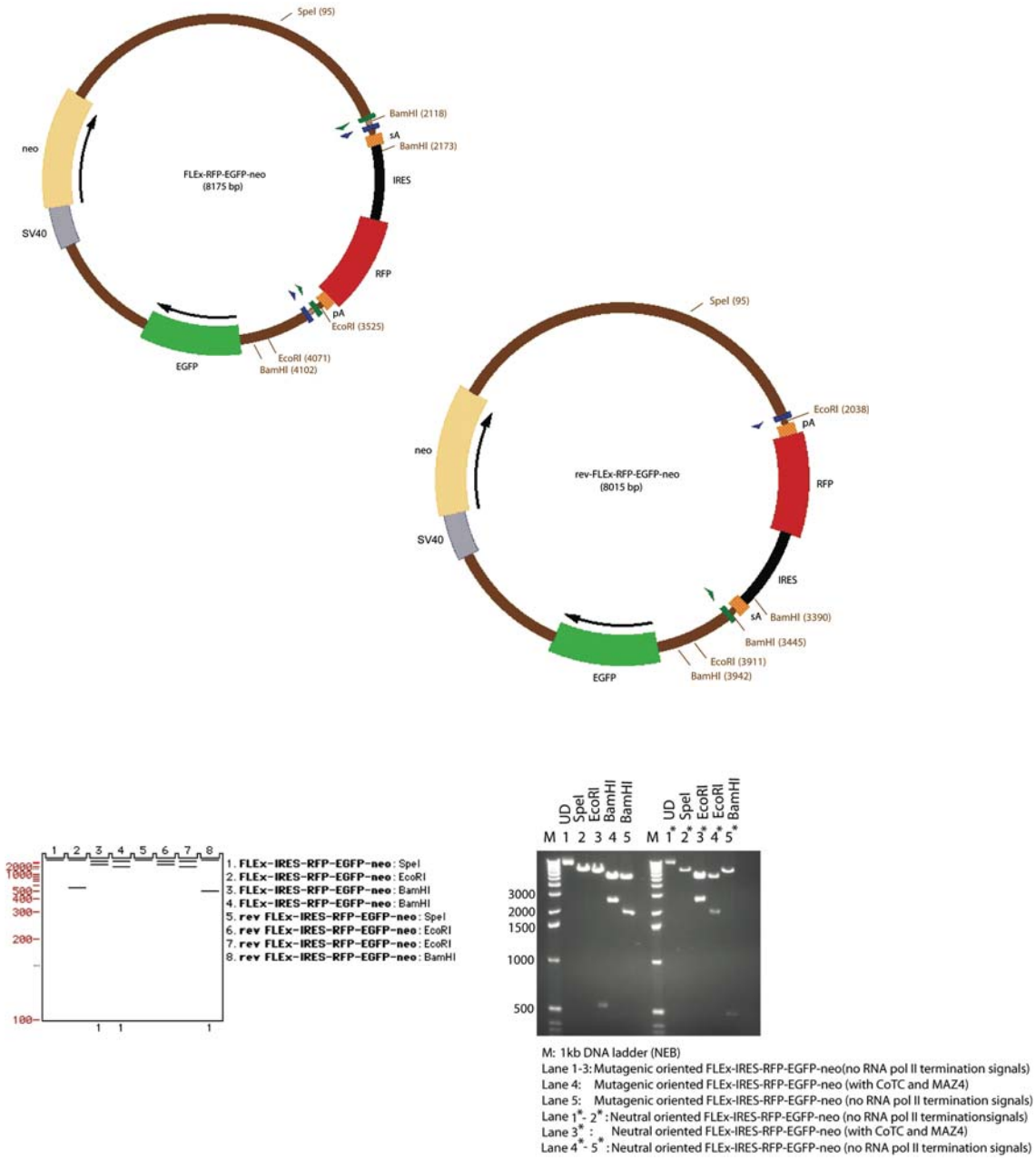


Figure 45 Map and agarose gel electrophoresis of analytical digests of pFLEx-IRES-RFP-EGFP-neo in both orientations containing no transcription termination signals before ES cell transfection

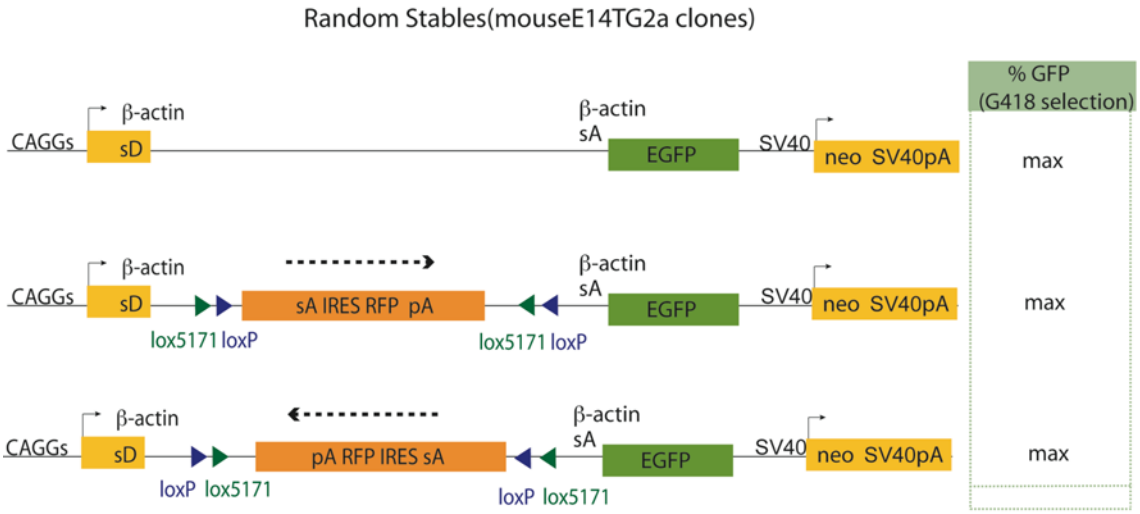


Figure 46 GFP assay results of random stables transfected with both orientations of the FLEx -EGFP-neo construct containing no pol II termination signals

2.1.2.1 Conclusion

Our results demonstrate that because cassettes lacking termination signals could not control expression of the relatively small EGFP gene, such a design should not be used for mammalian genes, which are more complex in structure and functionality. In conclusion, these signals are required for more sophisticated FLEx cassette design.

2.1.3 The best configuration of the FLEx cassette -Blasticidin constructs

There was no doubt about the functionality of pre-mRNA processing signals and the termination signal, CoTC, in the mutagenic oriented FLEx cassette. Following insertion of the cassette into the β-actin intronic region of the CAGGs-EGFP-neo vector, truncated EGFP mRNA was produced and high percentages of the random stables (E14TG2a clones) were characterized as being neomycine resistant and negative for GFP expression. Nevertheless, the reporter gene RFP was not functioning properly in recombinant clones. Because we needed a functional mammalian reporter gene to isolate recombinant E14TG2a clones among a heterogenic population in a reliable way during future gene trapping and targeting experiments, we selected BSD, an antibiotic resistant

gene. In neutral oriented second series of FLEx-EGFP-neo constructs, IRES-RFP was replaced with BSD under the control the strong hybrid of the SV40 and EM7 promoters using the Red-ET approach. All construct maps and restriction endonuclease patterns are summarized in figures 47, 48, and 49.

The restriction patterns of mutagenic oriented FLEx-BSD-synthetic pA containing both transcription termination signals are: NcoI, **703, 1135, 1518, 1809 and 2000 bp;** and, KpnI**, 641 and 6524 bp**. The restriction patterns of neutral oriented FLEx-BDS-synt pA are|: NcoI is **281, 693, 825, 1819, 2000 and 2074 bp (Figure 47)**. The restriction patterns of mutagenic oriented FLEx-BSD-synthetic pA containing only CoTC are: NcoI, **703, 1135,1364, 1809,2000 bp**; and, KpnI, **466 and 6545 bp**. The restriction patterns of the same construct in a anti-sense version (neutral) are: NcoI, **281, 703, 825, 1809, 1920, 2000 bp**; and, KpnI, **1908 and 5630 bp (Figure 48)**. The restriction patterns of mutagenic oriented FLEx-BSD-SV40 pA containing both CoTC and MAZ4 are: NcoI, **703, 1135, 1749, 1809, 2000 bp**; and, KpnI, **641 and 6755 bp**. The restriction patterns of neutral oriented FLEx-BSD-SV40 pA with both termination signals are: NcoI, **281, 703, 825, 1809,2000 and 2305 bp**; and, KpnI, **2150 and 5773 bp (Figure 49)**. For all BSD constructs, SpeI has a single site and was used for plasmid linearization before electroporation.

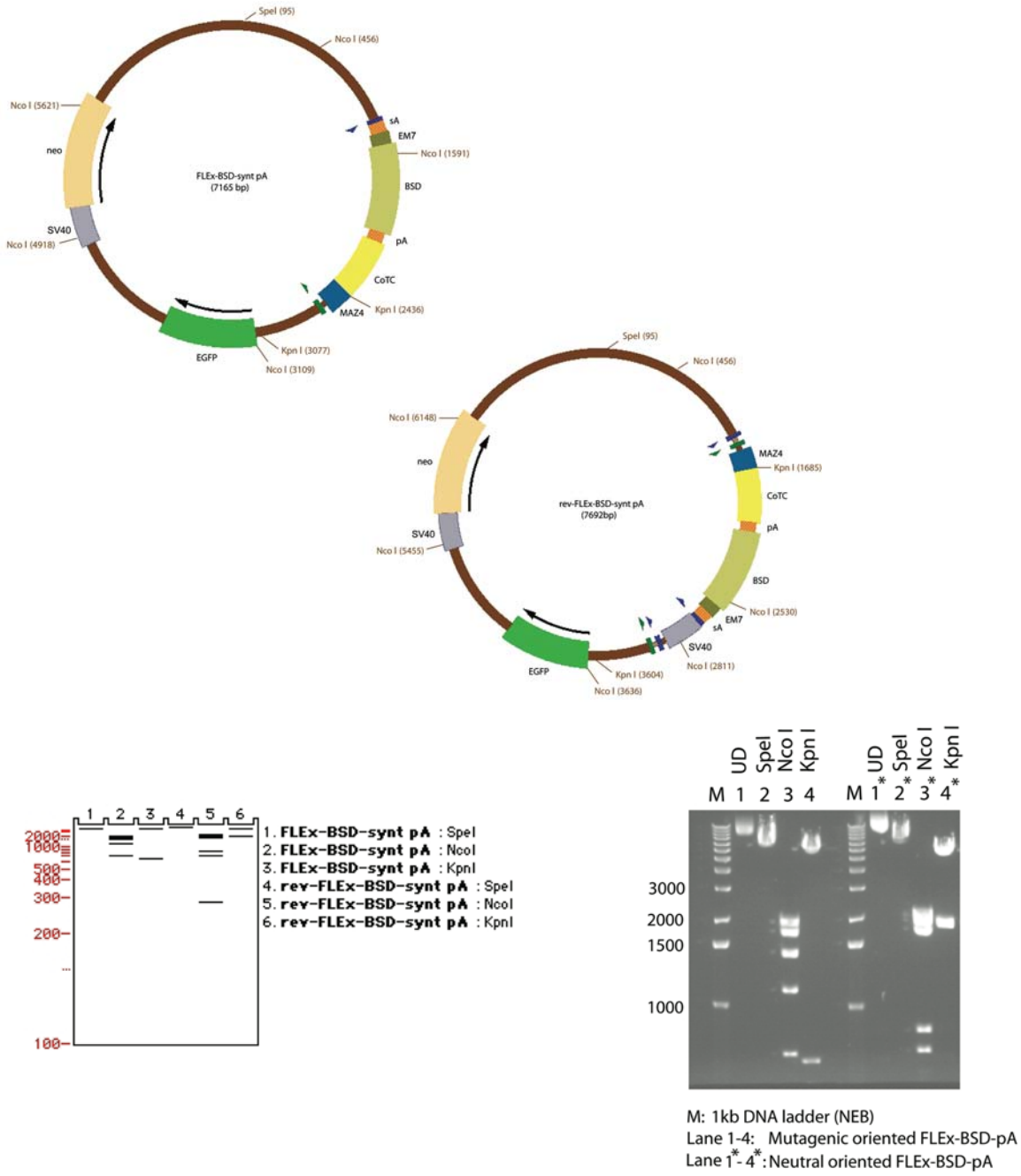


Figure 47 Map and agarose gel electrophoresis of analytical digests of pFLEx-BSD- synt pA -EGFP-neo in both orientations containing both transcription termination signals, CoTC and MAZ4, before ES cell transfection

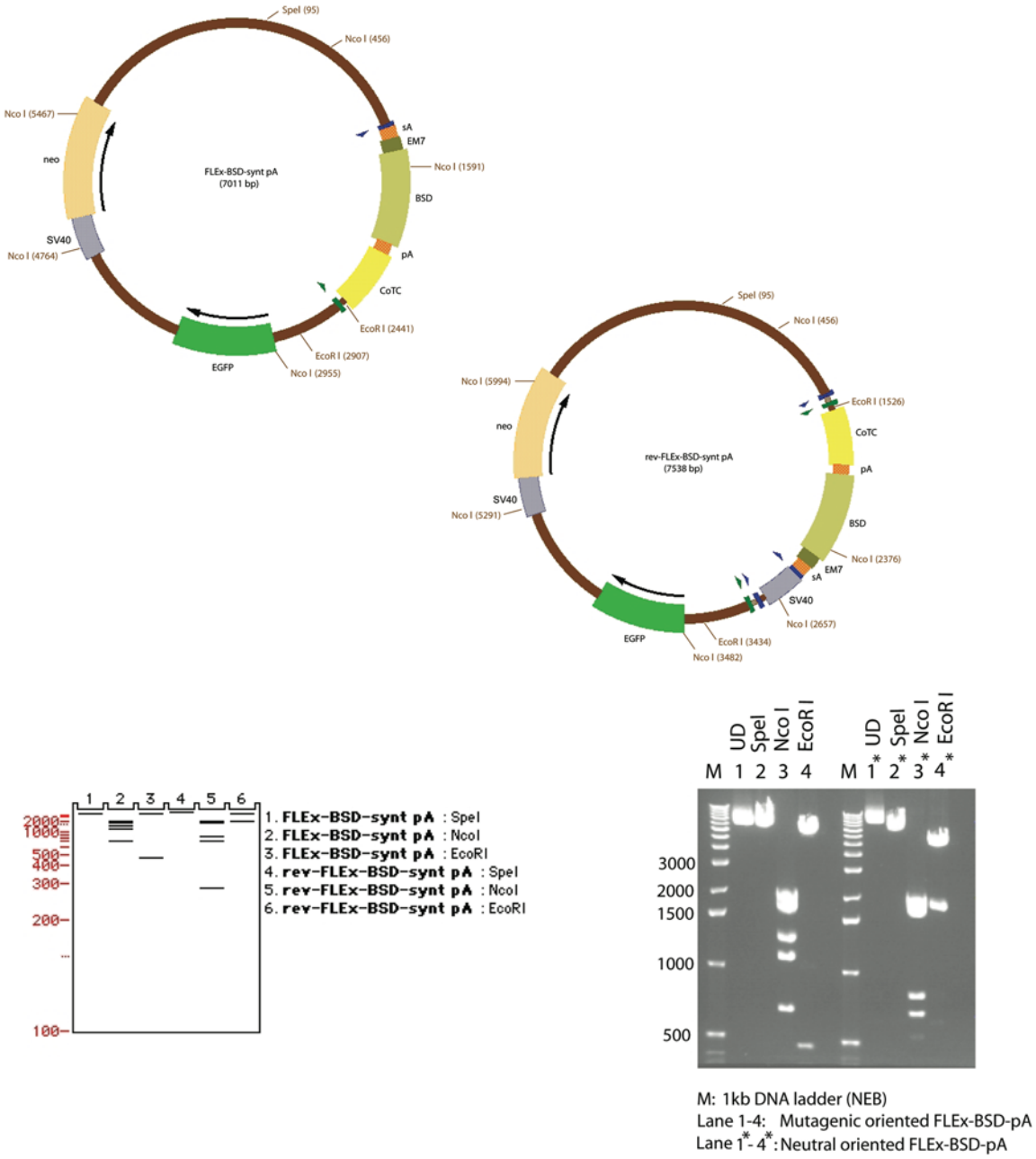


Figure 48 Map and agarose gel electrophoresis of analytical digests of pFLEx-BSD-synt pA -EGFP-neo in both orientations containing the single transcription termination signal,CoTC, before ES cell transfection

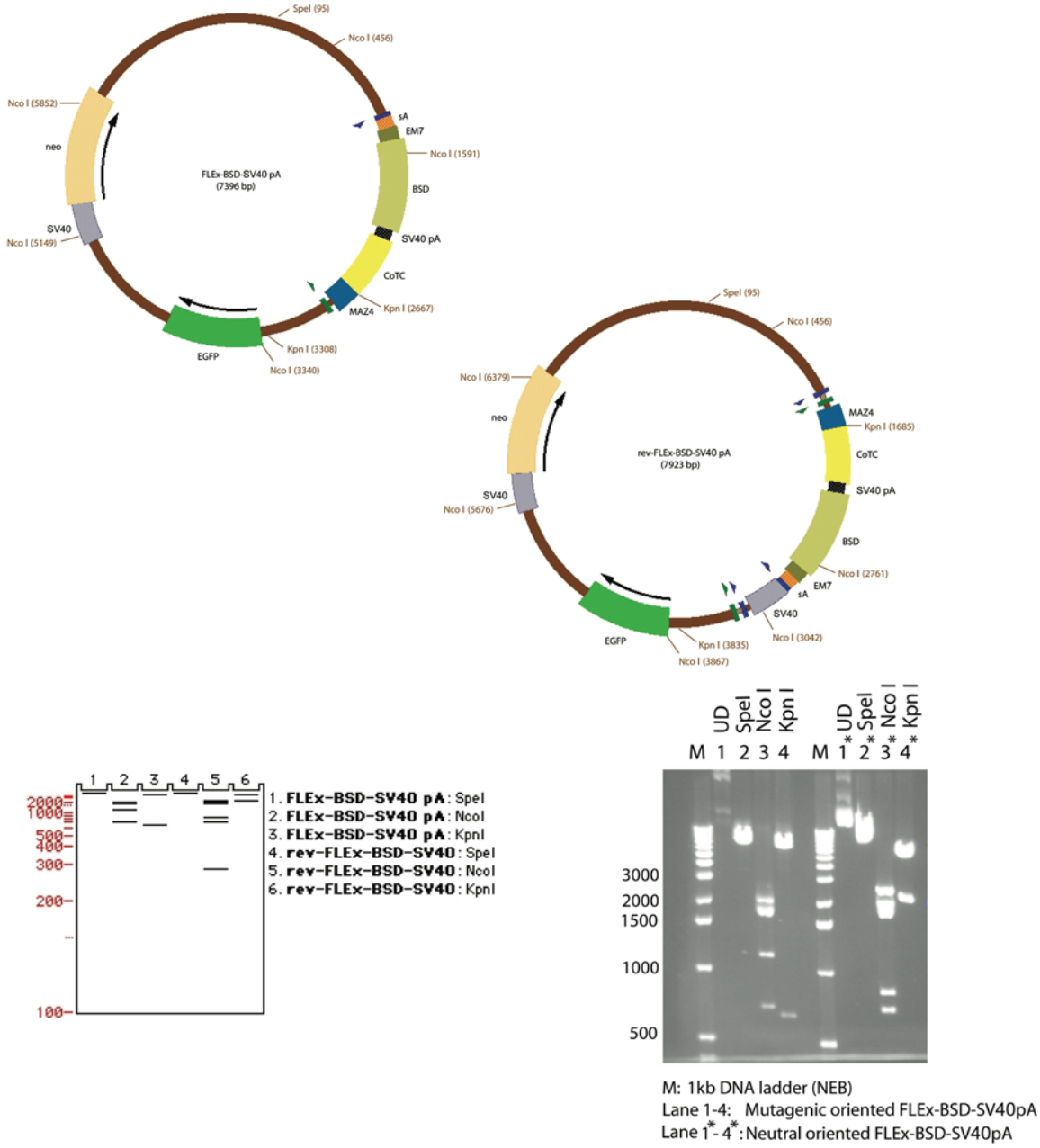


Figure 49 Map and agarose gel electrophoresis of analytical digests of pFLEx-BSD-SV40 pA -EGFP-neo in both orientations containing both transcription termination signals, CoTC and MAZ4, before ES cell transfection

In both FLEx cassette configurations, the SV40 promoter could drive BSD expression in ES cells. However, one disadvantage of using a strong promoter was its interference with both endogenous target gene and neighbouring gene expression. Another issue was the possibility that endogenous promoter competition could affect the number of clones produced by the recombinant ES cell lines. To prevent this problem, we inserted an extra loxP site in front of the SV40 promoter in the neutral oriented FLEx cassette. The aim was to use triple loxP recombination following cre induction, and remove the promoter in the complete knock-out mutation to prevent it from interfering.

In addition to these configuration changes, we thought to compare a strong pA element used in a lacZ expression reporter cassette made by Dr.Guiseppe Testa, with the synthetic one that appears in the FLEx gene trapping cassette. Because SV40 pA was bi-directional, we wanted to explore its contributions to both conditional and complete gene inactivation by generating FLEx. E14TG2a stable cell lines containing neutral oriented FLEx-EGFP-neo constructs that have either SV40 pA or another synthetic pA were generated. The FLEx construct containing only the MAZ4 pausing signal was excluded in this performance comparison. 24 hours after the electroporation, 3x10cm plates were subjected to G418 (0.20 μg/ml) and 3x10cm plates to BSD (5μg/ml). 14 days later GFP expression was assayed. As shown in figure 52, the neutral orientated cassette was reliable only in the presence of the synt pA signal. The percentage of GFP positive E14TG2a clones was nearly 50% lower with the same cassette in the presence of the synt pA element for G418 selection, and 71.5% lower for BSD selection. Only the random stables with neutral oriented FLEx-EGFP-neo containing only the CoTC termination signal, gave the same rate of GFP expression as clones subjected to both G418 and BSD selection **(Figure 50)**. This demonstrated that in this configuration the cassette functioned properly. Compared to G418 selection, BSD selection gave low numbers of clones. This was probably due to CAGGs and SV40 promoter interference.

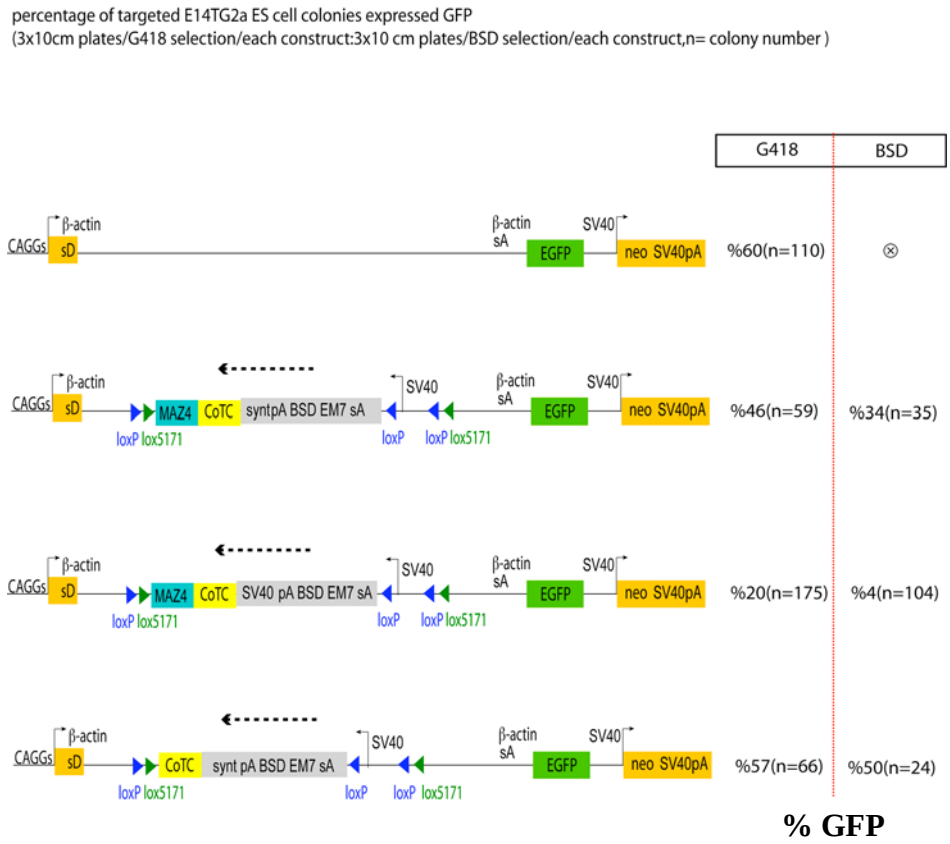


Figure 50 GFP assay results of neutral oriented third series of FLEx constructs transfected into mouse E14TG2a clones.

Fluorescent microscopy of random stable E14TG2a clones. Total clones (n) from each 3x10 cm plates were examined.

Cre recombination efficiency data (percentage of cell show no GFP expression, n = 50)			
Clones from BSD selection		Clones from G418 selection	
C1	55%	C4	71%
C2	64%	C5	85%
C3	95%	C6	60%

Table 5 Stable transfection of pCAGGS-Cre-puro into FLEx-CoTC-BSD-EGFP-neo E14TG2a clones

Cre recombination assay results of 6 different clones transfected with neutral oriented FLEx-BSD-EGFP-neo construct, half selected by BSD and the other half by G418 for 14 days before they were picked. The clones were characterized by total GFP expression before the electroporation of pCAGGs-Cre-puro plasmid. 24 hours after transfection, each clone was selected for 2 days in puromycin (1μg/ml). After 10 days without any selection, the statistics were done. The entire cassette inverted in almost all C3 derived, whereas C4 and C5 derived clones showed less efficiency.

In order to test the mutagenic version of the cassette, we performed Cre recombination to invert the neutral cassettes and isolated mutagenic FLEx in the prokaryotic system, *E.coli*. Then ES cell lines were electroporated with FLEx-EGFP-neo containing either both transcription termination signals or CoTC alone. All clones produced by these cell lines were then checked for GFP expression. The cassette in the CAGG promoter direction created a strong mutation, completely shutting off GFP expression in random stables **(Figure 51)**. In mutagenic oriented constructs, expression of the selection marker BSD was under the control of the CAGGs promoter because the SV40 promoter was removed following Cre recombination. This prevented SV40 and CAGGs competition. Compared to the neutral FLEx-EGFP-neo E14TG2a cell lines, the mutagenic FLEx-EGFP-neo cell lines tended to generate a higher percentage of healthy recombinant clones in which GFP transcription terminated completely.

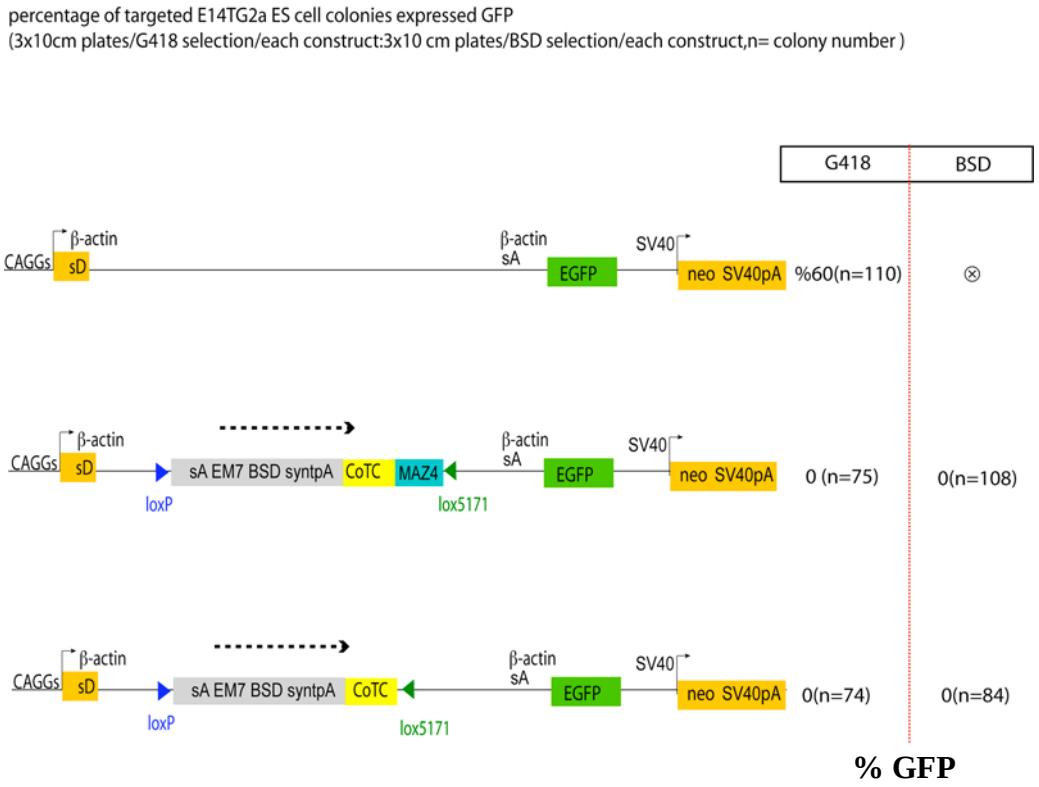


Figure 51 GFP assay results of mutagenic oriented third series of FLEx-EGFP-neo mouse E14TG2a clones

To confirm the absence of EGFP expression, proteins were extracted from some mutagenic oriented FLEx-BSD-EGFP-neo clones and western blotting was performed. Extract from pcDNA6V5-His-SV40-BSD transfected clones was used as a negative control and extract from pCAGGs-EGFP-neo empty vector transfected clonse was used as a positive control. All the random stables showed no trace of an EGFP specific band **(Figure 52)**.

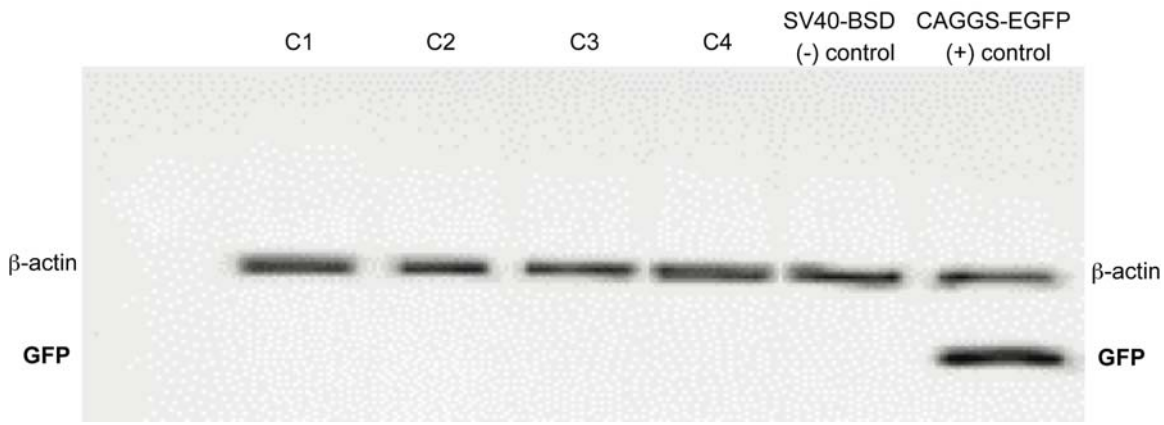


Figure 52 No GFP expression was detected in mutagenic oriented FLEx-BSD-CoTC-EGFP-neo clones

Protein extracts from random stables (E14Tg2a clones) and two control clones probed with a specific mouse polyclonal antibody recognizing the N terminus of recombinant GFP protein (EGFP), 29 kDA monomers with 265 a.a. No specific EGFP bands were detected in random stables (C1and C2 G418 selected,C3 and C4 BSD selected) and the (-) control clone transfected with pcDNA6V5-His-SV40-BSD. All extracts were treated with goat β-actin polyclonal antibody that recognizes the 45kDA protein to demonstrate equal amount of proteins were assayed.

2.1.3.1 Conclusion

While constructing FLEx-EGFP-neo containing a BSD marker, we found that a unidirectional synt pA element in combination with CoTC, a strong termination signal, contributed substantially to cassette mutagenecity as reflected by the loss of the downstream EGFP transcription. The anti-sense orientation of these signals did not interfere with GFP expression. Recombinant clones, selected by either G418 or BSD selection, displayed a stable pattern of GFP expression, showing no variability in the GFP expression rate. Also random stables (E14TG2a clones) showed higher levels of cassette inversion following Cre induction. Without any doubt, the gene trapping cassette, FLEx in this configuration will enable us to achieve our aim of creating a targeted trapping strategy that is applicable to the majority of mouse genes.

2.2 Multipurpose Jarid1c Allele

In the second part of the project, we aimed to conduct a targeted gene trapping strategy using the best variation of the FLEx cassette **(section1.5)**. First, we wanted to insert the neutral orientated FLEx cassette into the first intron of a mouse gene to establish a multipurpose allele of this gene**.** To simplify this tactic, the mouse Jarid1c gene on the X chromosome was selected **(Figure 53)** because this gene is in single copy in mouse E14Tg2a ES cells derived from 129P2/OlaHsd males. This meant that as opposed to autosomal gene targeting, only one-step gene targeting was required for Jarid1c.

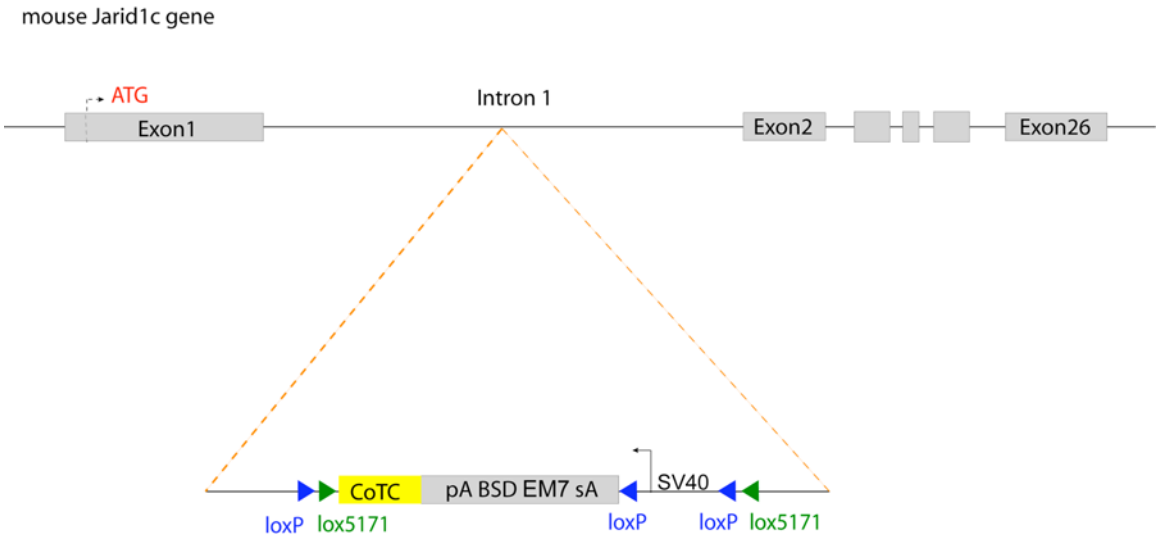


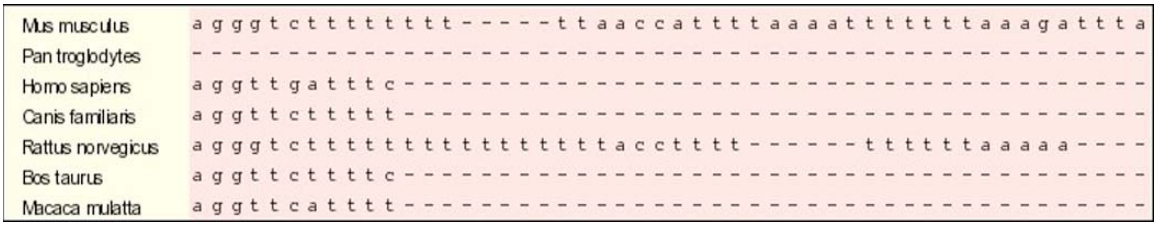
Figure 53 General scheme of FLEx cassette insertion into the mouse Jarid1c gene

2.2.1 Construction of conditional targeting Jarid1c vector

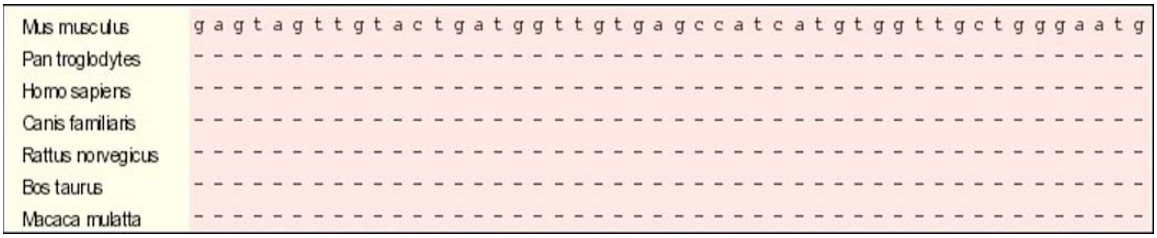
From the in-silico mouse genome sequence database (ENSEMBL), the Jarid1c gene was identified and mapped. It spans 40.5 kb of genomic sequence on Chromosome Xp11.2. It is composed of 26 exons. The annotated BAC clones in ENSEMBL were RP23 and RP24 series from the C57/Bl6 strain and CH29-179C2. Because no 129 strain BAC clone was listed, targeting vector construction was conducted with the RP23-147K15 clone.

The target intronic region that we aimed to insert the neutral oriented FLEx gene trapping cassette into is 3.9 kb. Before starting construction of the conditional allele, it was important to map potential Jarid1c transcription regulatory regions of the target site. Intron 1 of Jarid1c lays on the X chromosome between the coordinates 147.573.348-147.578.348 bps. Genomic DNA alignment of the intron I region between these coordinates in mouse was done against 6 mammals, *Pan troglodytes*, *Homo sapiens*, *Canis familiaris*, *Rattus norvegicus*, *Bos taurus* and *Macaca mulata*. The unconserved region between 147.577.072-147.577.167 was selected for cassette insertion **(Figure 54).**

(A) Alignment result of region 147.577.011-147.577.059



(B) Alignment result of region 147.577.072-147.577.119



(C) Alignment result of region 147.577.120-147.577.167

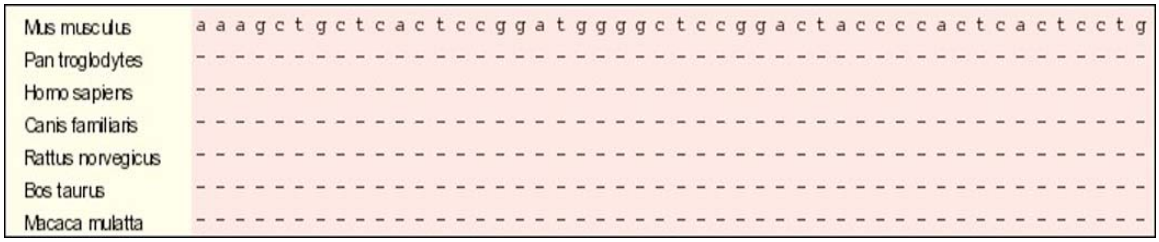


Figure 54 Partial alignment profile of intron 1 of Jarid1c

Part of the intronic region between 147.577.011-059 is conserved between the 6 mammals shown except *Pan troglodytes* (Figure A). Although it is not shown, the intronic region between 147.574.427-147.577.056 showed this conservation profile with some single nucleotide variations. Starting with the region 147.577.056-147.578.348 in *Mus musculus*, there was no conservation for the rest of the 6 mammals (unconserved regions). Figure B and C show part of these unconserved regions, and the cassette was inserted between 147.577.119-147.577.120.

The cassette was inserted with one-step Red-ET recombination into the unconserved region between 147.577.119-147.577.120. 48 bp sequences between

147.577.072-119 served as the homology arm portion of the forward PCR primer, and 48 bp sequences between 147.577.120-167 served as homology arm portion of the reverse PCR primer. This region was targeted with PCR product containing the FLEx cassette flanked by 48 bp homology arms on each side. After introduction of a pSC101 plasmid containing an arabinose inducible Red-ET recombination system into the 147K15 Jarid1c BAC containing host, tetracycline (3μg/ml) resistant BAC colonies (DH10B strain) were selected. Then, after induction of Red α,β,γ expression, the PCR product was electroporated into the BAC colony carrying pSC101 plasmid. Through Red-ET recombination, the first intronic region of Jarid1c with neutral oriented FLEx gene trapping cassette was modified. Positive colonies were selected with BSD (5μg/ml) and CM (15μg/ml). After collecting the candidate recombinant bacterial clones, mini-prep plasmid DNA analysis was done with PshA1. This enzyme has 13 sites on the BAC backbone and one site in the FLEx cassette. The recombinant 2.6 kb band could be easily identified among the non-recombinant colonies **(Figure 55)**.

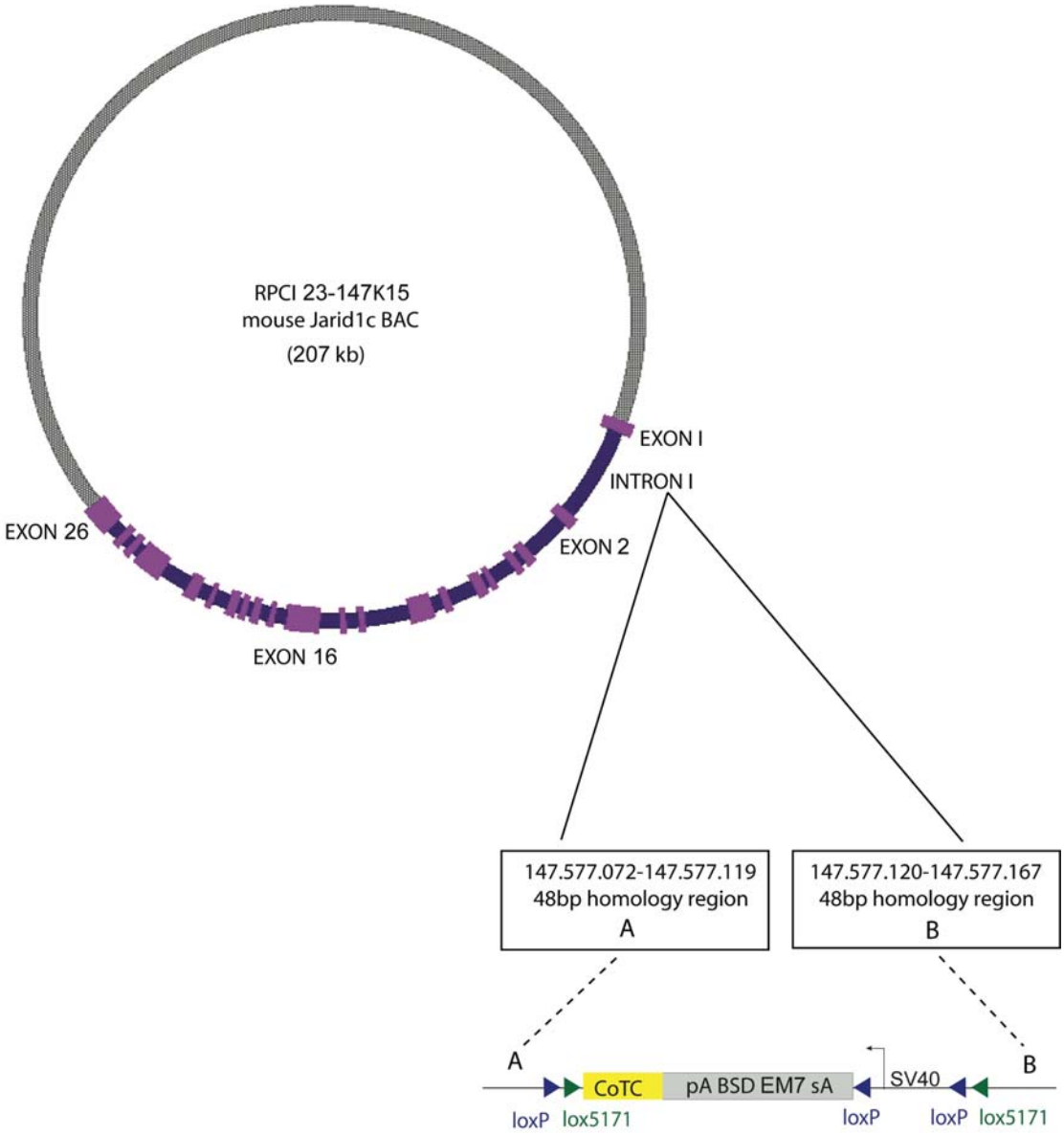


Figure 55 (A) The modification of the wild-type Jarid1c BAC DNA(147K15)

Insertion of the FLEx cassette to the first intronic region of the Jarid1c BAC (207 kb). The coordinates of the 48bp oligonucleotide homology arms A and B correspond to the positions of oligonucleotides on Xp11.2.

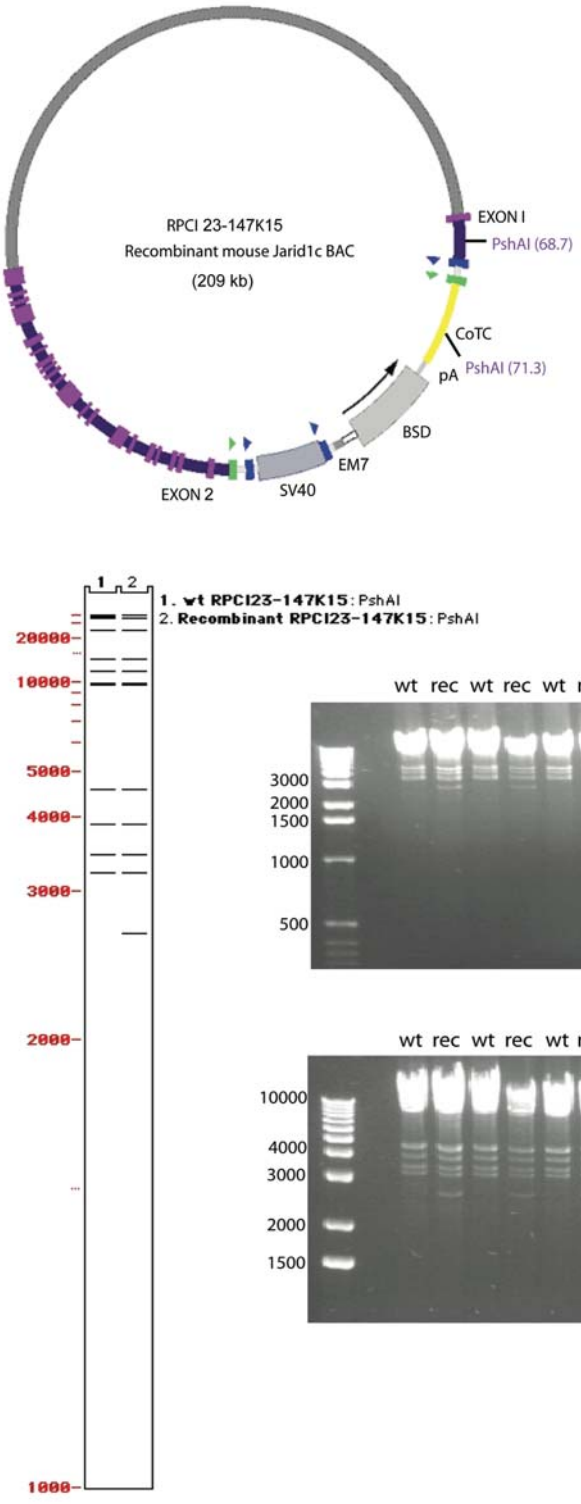
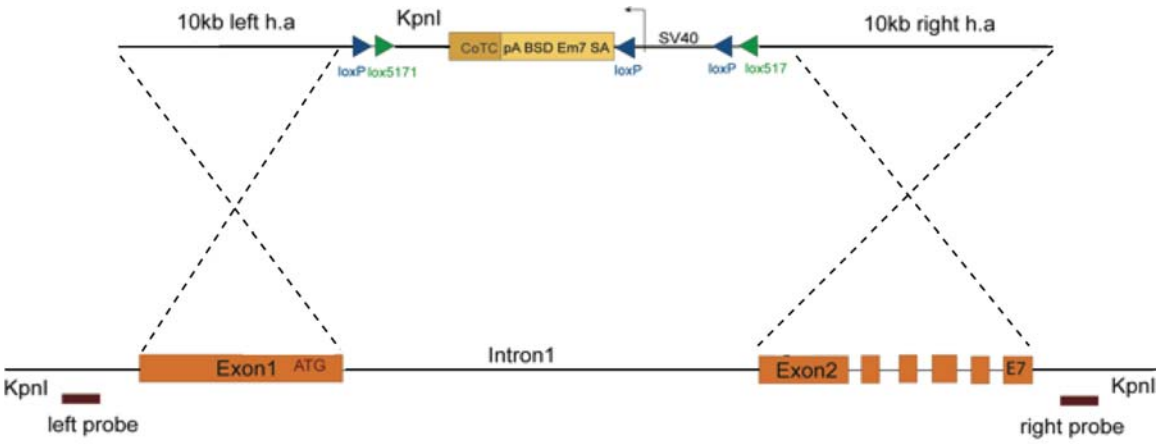


Figure 55 (B) Restriction analysis of Jarid1c BAC DNA after PshAI digestion

Figure 55 (B) shows the restriction analysis of recombinant BAC DNA after the cassette was inserted into the intron I target region. Recombinant clones were detected by PshAI restriction analysis. This enzyme has 13 sites on the BAC DNA backbone, one critical position is at 68.7 and the other site was positioned at 71.3 in the FLEx gene trapping cassette. The restriction pattern of the recombinant band was 2.6 kb.

The next step after finding the correct recombinant Jarid1c BAC was to generate a targeting construct for homologous recombination in a mammalian system, mouse ES cells. In order to enhance homologous recombination frequency, because a non-isogenic DNA source was used in this project (C57/Bl6), the homology arms of the targeting vector were increased. Using a convenient southern blot strategy, we designed the 21.9 kb targeting vector with homology arms containing 10 kb of the first Jarid 1c exon, including the translational start site, (left arm) and 10 kb of exon 2-7 (right arm). There was more DNA in the arms than in the floxed region in order to enhance Cre recombination at the target. The 30 kb Jarid1c transcript sequence (ENSMUST 00000082177) was aligned, using coverage information from Celera strains **(information from Dr.Daniel Rios, EBI)**, against 129S1/SvImJ and 129X1/SvJ. As shown in figure 39, there was we were unable to get sequence information for 129S1/SvImJ that covered the left arm homology region from the C57/Bl6 strain. However, 2.7 kb of sequence information were available for the 129X1/SvJ strain and these were 100% homologous to the C57/Bl6 sequence information. In the case of the right arm, no sequence information from the 129S1/SvImJ strain was available in this region. For the 129X1/Sv strain, the 3.9 kb coverage information available was homologous to the region from C57/Bl6 **(Figure 56)**.

(A) Design of Jarid1c targeting vector based on Southern blot strategy



(B) Alignment results of 24kb transcript from C57/Bl6 against two 129 substrains

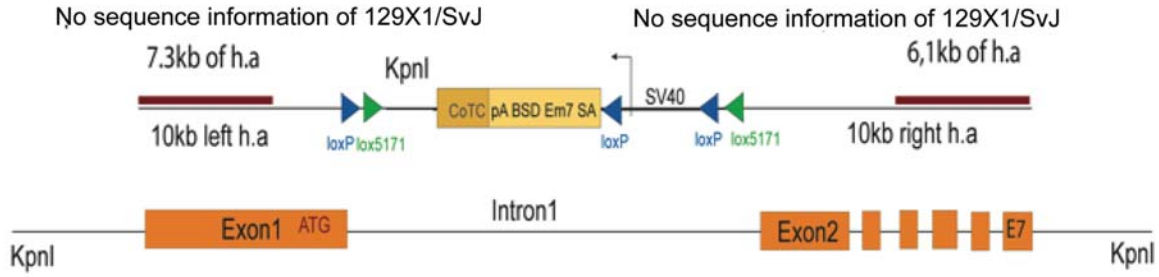


Figure 56 (A) For the convenient southern blot analysis of recombinant and wild type Jarid1c alleles, we selected the KpnI restriction site. Depending on the presence of these restriction sites in the wild type Jarid1c allele, and on the specific probe coordinates, we determined the suitable regions for the 10 kb homology left and 10 kb homology right arm for the vector.
(B) 30 kb Jarid1c transcript sequence from C57/Bl6 was aligned against two 129 substrains, 129S1/SvImJ(not shown) and 129X1/SvJ. For the homology arm regions, no sequence data was obtained from the Celera 129S1/SvImJ strain. As for the 129X1/SvJ strain,7.3 kb sequence information for left, and 6.1 kb information for the right arm were missing. On the other hand, 2.7 kb left region and 3.9 kb right region sequence information of 129X1/SvJ showed 100% homology to the C57/Bl6 sequence information.

Subcloning of the 21.9 kb targeting the Jarid1c region into the 3.9 kb pACYC177 was the last prokaryotic DNA engineering issue in this project. For the Red/ET recombineering to generate the Jarid1c targeting construct, a 3.9 kb minimal vector was amplified with specific primers that flanked the homology arm which corresponds to the targeting vector and the restriction enzyme NotI, using pACYC177 as a template. Through homologous recombination in the prokaryotic system, the Jarid1c targeting backbone was cloned to a minimal vector version of pACYC177 (**Figure 57)**. The candidate recombinant clones were picked using BSD (5μg/ml) and ampicillin (100μg/ml) selection, and restriction enzyme analysis and sequencing was then used to find the correct recombinant clone (**Figure 58)**.

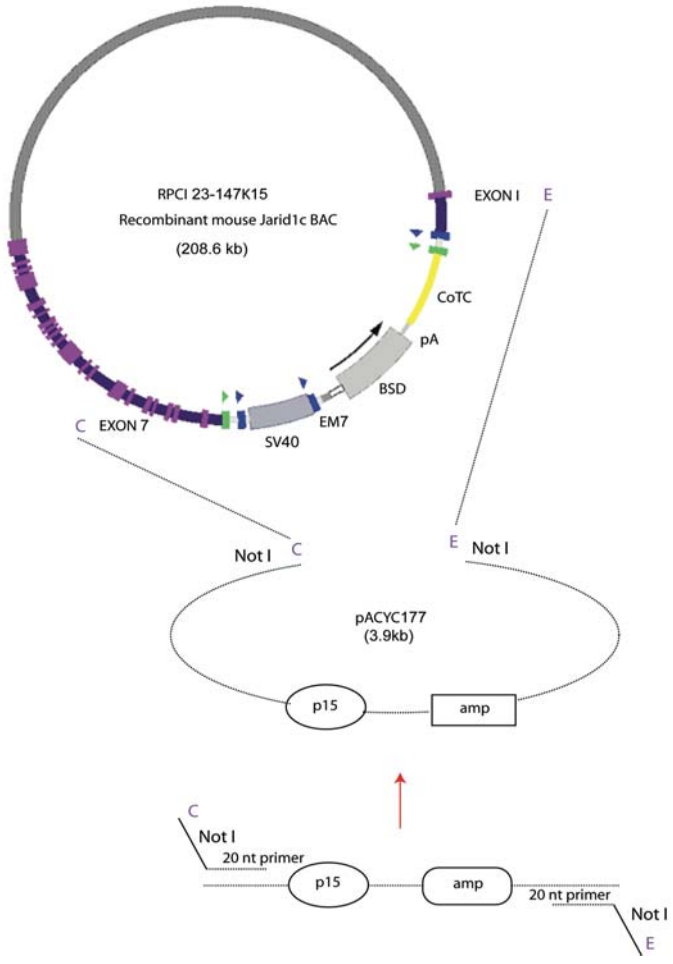


Figure 57 Subcloning the targeting backbone into the linear pACYC 177

To generate the targeting vector, pACYC177 derived minimal vector was amplified with 5' C (homology arm to BAC vector) NotI A (20 nt primer of pACYC177), and 3' NotI and E (homology arm to BAC vector). Via Red-ET recombination, the targeting BAC region ExonI-Exon7 was cloned into the pACYC177

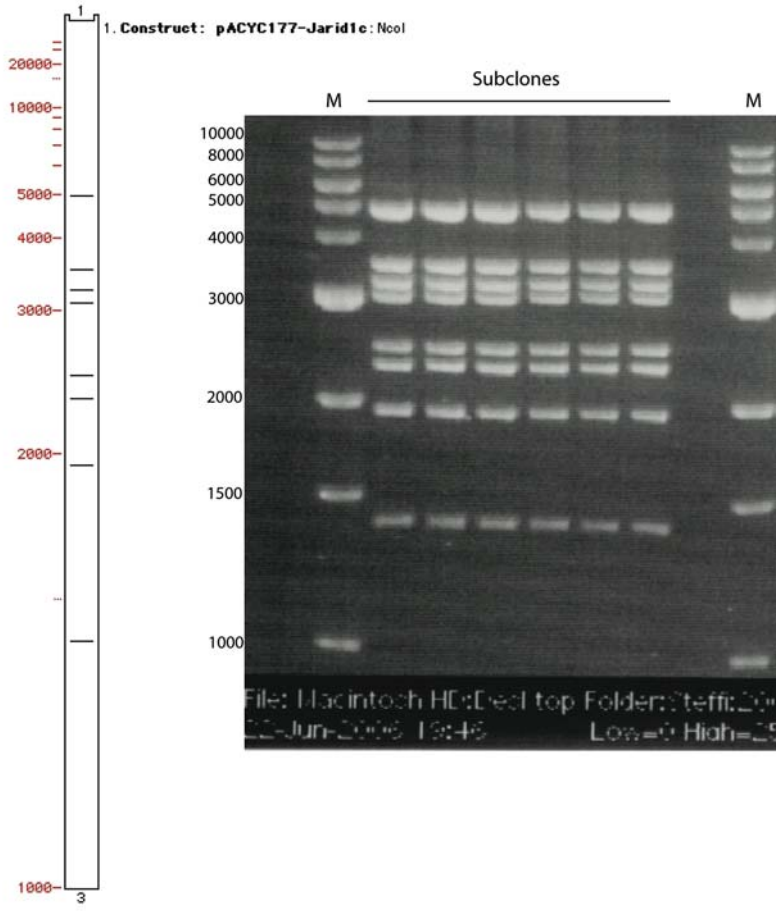


Figure 58 (A) Verification of pACYC-Jarid1c subclones

Restriction analysis of pACYC177-Jarid1c subclones after NcoI digestion. The restriction pattern of pACYC177-Jarid1c for NcoI is 281bp, 473 bp and 631 bp (not shown), 1397bp, 1949bp, 2295bp, 2440bp, 3062bp, 3217bp, 3484 bp and 4964 bp. Among these different subclones, the correct one was verified with sequencing analysis of the 1 kb left and 1 kb right mammalian homology arm of the Jarid1c backbone.

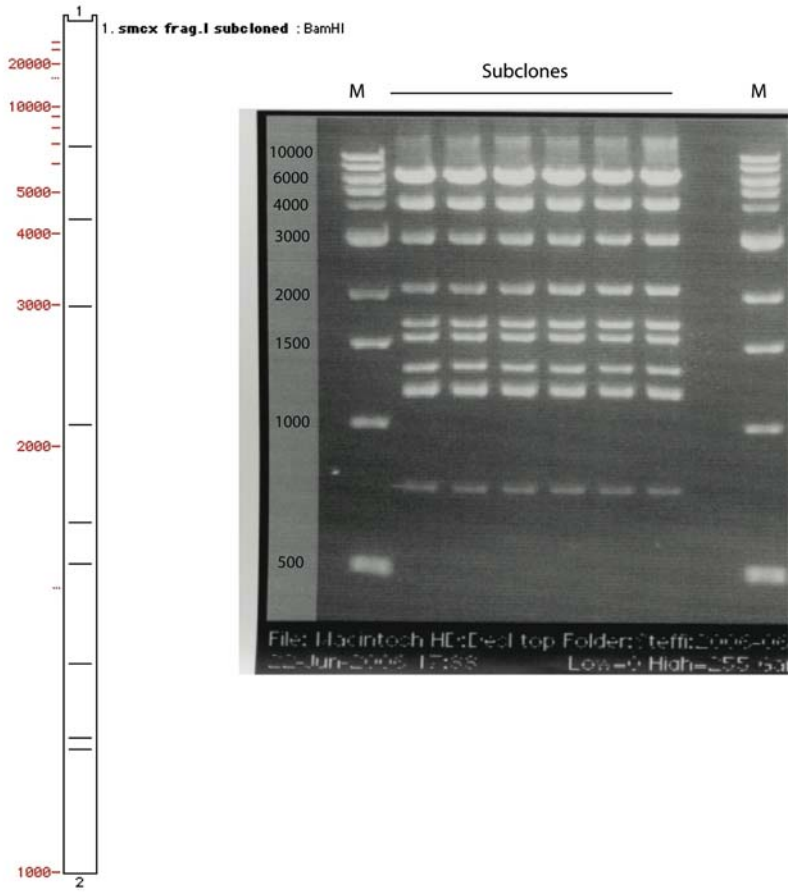
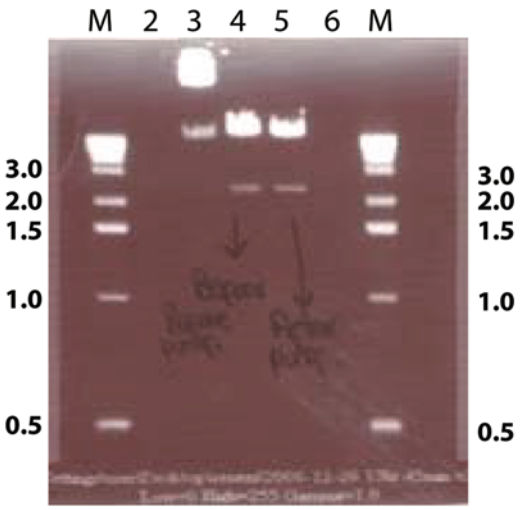


Figure 58 (B) Verification of pACYC-Jarid1c sublones

Restriction analysis of pACYC177-Jarid1c subclones after BamHI digestion. The restriction pattern of pACYC177-Jarid1c for BamHI is 348(not shown), 758 bp, 1168bp, 1186bp, 1324bp, 1566bp, 1698bp, 2105bp, 2986 bp, 4291 bp and 6763 bp. Among these different subclones, the correct one was verified with sequencing analysis of the 1 kb left and 1 kb right mammalian homology arm of the Jarid1c backbone.

To target Jarid1c through homologous recombination, the targeting vector was linearized with NotI and electroporated into E14TG2a cells. 3 rounds of electroporation were performed **(Figure 59)**. 24 hours after the electroporation, BSD selection (5μg/ml) was performed for 14 days, then undifferentiated clones were picked from the 96 well plates and later expanded to 24 well plates.



30 μg DNA/ in 30 μl TE electroporated into $8x10^6$ E14TG2a cells

AUGUST 2006 : 11 clones/ BSD selection(5μg/ml)

OCTOBER 2006 : 87 clones/ BSD selection(4 μg/ml)

30 μg DNA/ in 30 μl H_2O electroporated into $8x10^6$ E14TG2a cells

DECEMBER 2006 : 16 clones/ BSD selection(5 μg/ml)

Figure 59 Jarid1c gene targeting experiments

The Jarid1c targeting vector was linearized with NotI. NotI digestion of the pACYC177-Jarid1c subclone gave the 2.204 kb the pACYC177 vector backbone, and the 21.989kb Jarid1c targeting backbone. All DNAs were purified using phenol-chloroform. Lane 3 is undigested pACYC177-Jarid1c subclone, lane 4 is digested pACYC177-Jarid1c vector before DNA purification and lane 5 is the digested pACYC177-Jarid1c subclone used for electroporation of the EScells.

The 114 candidate recombinant clones from the 3 electroporations were screened by southern blotting **(Figure 60)**. Southern blotting probes were designed with a special bioinformatics software tool, CENSOR (www.girinst.org/censor, Jurka J. et al), that screened query sequences against a reference collection of repeats and masked homologous regions with masking symbols, as well as generated a report classifiying all found repeats. In this project, the candidate left and right Jarid1c probe regions were analyzed with the vertebrate and mammalian reference collection of repeats and masks. This showed that the candidate right probe regions were not suitable and that the left probe regions had a couple of target sites that did not show repeats. One of the candidate left probe regions was selected and the probe was designed from it. Additionally, the probe was blasted against the genomic sequences of Jarid1d, a homologue of Jarid1c. The blast results indicated that the probe could only recognize the X chromosome region of Jarid1c gene.

The last targeting experiment was conducted with feeder dependent mouse ES cells, IB10 from the E14 origin. This showed 50 clones were the product of illegitimate recombination instead of homologous recombination.

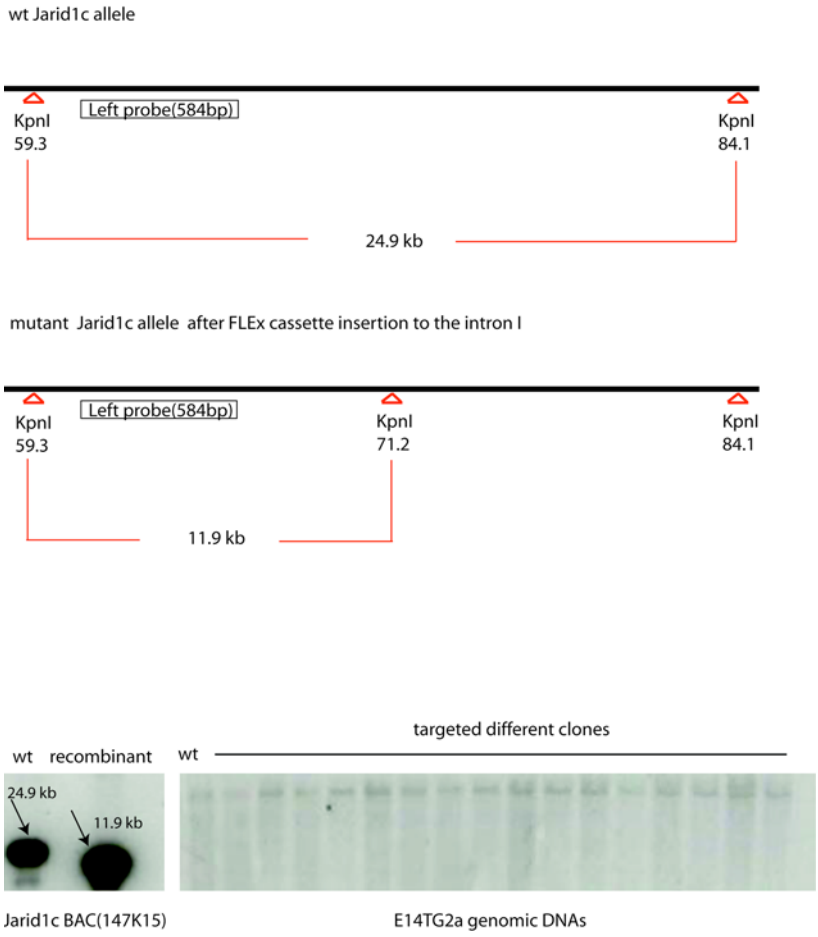


Figure 60 Southern blot analysis of the Jarid1c allele

Representation of southern blotting of wild-type and mutant Jarid1c alleles (upper). BAC or genomic DNA from were digested with KpnI enzyme and then hybridized with the specific probe (584 bp). The wild type allele was identified by a 24.9 kb fragment and the mutant allele was identified by a 11.9 kb fragment. Recombinant and wild-type allele bands were observed (left). The cassette was correctly inserted into the first intronic region of the Jarid1c gene. The Southern blot was probed with the left probe of the KpnI digested genomic DNA from wild type and different targeted clones confirmed that the gene could not be targeted; all BSD resistant clones were the product of illegitimate recombination.

3. DISCUSSION

3.1 Efficient Trapping of the EGFP Transcript by A Generic FLEx Cassette

Gene expression is not solely regulated by RNA polymerase II transcription efficiency or RNA polymerase II dependent transcription termination. Other factors have a significant effect on gene expression levels and control gene transcription in coordination with RNA polymerase II transcription. Chromatin state can also seriously impact transcription and pre-mRNA processing events are essential to regulation of gene expression. In summary, the state of the chromatin where a gene is located and the interaction between RNA polymerase II transcription and pre-mRNA processing events are prominent features that coordinate to positively or negatively influence gene expression.

In this PhD project, FLEx was engineered to control the expression of a wide range of endogenous mouse genes by affecting RNA processing events. We found that efficient splicing and polyadenylation were the two main criteria that need to be fulfilled by the mutagenic oriented FLEx gene trap cassette. When the cassette was inserted into the first intronic region of any mouse gene in the same direction of its promoter, the cassette served as a synthetic terminal exon because the sA and pA signals of the cassette were recognized by the pre-mRNA processing machinery. These exogenous pre-mRNA processing signals prevented recognition of the endogenous sA and pA. This prevented the trapped gene from being transcribed downstream of the cassette insertion region. It is now clear from the work presented here that when building a gene trapping cassette to create RNA transcript level mutations it is important to consider the contribution of splicing and polyadenylation signals. Because these signals are so important for efficient mRNA processing both their structure and proximity to each other should be verified.

Constitutive or alternative intron splicing is very critical for co-transcriptional events that ensure correct protein production via reliable mRNA processing. In a coordinated

manner, the splicing machinery elements recognize and load on to specific sequences at the 5^I (splicing donor) and 3^I (splicing acceptor) ends of introns. The splicing mechanism is very complex in higher eukaryotes, but in general it starts with the formation of a of complex ribonucleoproteins, called the spliceosome. In higher eukaryotes this mechanism requires more protein-protein interactions. The formation of this complex at specific splice site regions is enhanced by recognition of 5^I splice site through U1 snRNP, followed by recognition of the branch point by the branch point binding protein SF1 and the U2AF helper protein, and recognition of the 3^I splice site by U2 snRNP. snRNPs U4-U6, excluding U3, are also involved in pre-mRNA splicing.

Vertebrate splicing complexity is greater in comparison to other eukaryotes because vetebrate genes are more intricate. They often have very large introns and multiple small exons that average at 137 nucleotides **(9)**. This exon-intron architecture makes exon/intron boundries more difficult to identify. Therefore spliceosome assembly requires a well-defined and strong recognition region. Splicing efficiency is determined by two factors (1) distance between specific splicing signals including the splicing donor, branch sequences, polyprimidine sites and the sA, and (2) their relative strength which is determined by the presence of an uninterrupted polypyrimidine tract. Based on the exon definition theory, exons are first recognized as a unit during the early assembly by factors that recognize and bind to the 3^I end of intron (SD), and then the search for the splice acceptor begins. This suggests that direct communication occurs between the factors that recognize splice donor and acceptor sites **(9,58,75)**.

To maintain this precise communication and prevent exon skipping, the minimum distance between the splice donor and splice acceptor sites is required to be about 100 nucleotides in length **(personnel communication with Juan Valcarcel)**. In this study, the FLEx cassette contained a synthetic splice site that was composed of a consensus branch site, a pyrimidine rich region and splice acceptor. The proximity of the branch site was 887 bp to the downstream splicing donor and 27 bp to the upstream acceptor region, the optimal distances **(111)**. In addition to the appropriate distance between splice recognition sites, the uninterrupted polyprimidine tract was an important component of

this cassette as it prevented the splicing machinery from skipping the synthetic terminal exon generated by FLEx in the intronic region of pCAGGS-EGFP-Neo.

Another important factor affecting transcription is proper polyadenylation. A pA tail is found on nearly every processed mature mRNAs in vertebrates. It functions to stabilize mRNA, aids its transport from nucleus to cytoplasm, and promotes translation efficiency. Precise pA signals influence the recognition of the polyadenylation machinery elements and enhance efficient 3^{I} end formation of the newly synthesized pre-mRNA. These signals are composed of a consensus sequence, usually A residues at the 5^{I} site and GT/T rich sequences at the 3^{I} site. The distance between these two sequences determines the strength of pA signal and thus robust mRNA end formation. The approximate distance between these two cleavage and pA addition signals should be 20-40 nucleotides in length, a criteria fulfilled by the FLEx cassette synthetic pAsignal **(18,19)**. These sequences are recognized by specific proteins; the CPSF factor interacts via its large unit with the A residues and the CstF factor recognizes and binds to the GT/T rich region. These factors involved in polyadenylation are composed of several subunits.

Many experiments conducted on transcription and RNA processing machinery in yeast and mammalian cells have revealed a connection between these two events. This indicated that these factors associated with pre-mRNA 3^{I} end processing, cleavage and poly(A) tract formation, interact through their subunits with constituents of other processing events such as splicing **(9,58)**. In particular, important support for this hypothesis has come from the isolation of factors involved in both polyadenylation and splice regulation, as well as in the recognition of the 5^{I} splicing site of the exon. The U1A protein is well known for playing a role in the regulation of these two events. It was found that SR proteins, key regulators of the 5^{I} splice site selection have a connection with polyadenylation. They have rich arginine - serine domains (RS) and through these domains they interact with U1A and U2A snRNPs and other proteins composed of this RS units such as the polyadenylation cleavage factor, CFI. These factors bind the enhancer sequence of the exons or alternatively directly bind to the exon splice site. In addition to promoting recognition of the exon by the splicing machinery, they activate polyadenylation cleavage **(57)**. One study of calcitonin pre-mRNA processing claimed

that polyadenylation possibly communicates with splicing through SR proteins and that this may enhance polyadenylation. How is this possible? According to a study in human thyroid cells, the SRp20 factor binds to an enhancer region consisting of the 5^I splice site of exon 4 preceded by a pyrimidine tract. The 5^I splice site is recognized by U1 RNA (one of the small nuclear proteins) and other splicing machinery members. The polyadenylation cleavage factor CFI also consists of serine and arginine dipeptides. Interaction of these two SR proteins enhances polyadenylation at the 5^I splice sites. Mature mRNA structure of calcitonin in thyroid cells includes exon1-4, although the primary transcript has 6 exons. In this theory, this suggests that SR proteins can influence pre-mRNA processing at the level of polyadenylation and generally participate in the recognition of terminal exon, referred to in this PhD study as "last exon recognition". Although all these observations were built on assumptions, they are highly suggestive of cross-talk between splicing and polyadenylation **(57,83)**.

Given the above data, we think that communication between elements requires that they are located in regions that enable them to recognize their binding sites, and that promote efficient splicing and polyadenylation. In this scheme, certain distance requirement between the splicing signals and poly A signal recognition factors must be observed. Based on the exon definition theory, terminal exons, namely both first and last exons in vertebrates, and internal exons are recognized by different mechanisms. The recognition of the terminal exon, referred to here as the last exon, is dependent on an interaction between the splice acceptor and polyadenylation factors U1 and U2snRNP at a 3^I splice site recognizing factors U2AF and SC35 **(9,75)**. In order to promote this interaction, at least 100 nucleotides length distance should be generated between the sA and pA in the FLEx gene trapping cassette, similar to the proximity required between splicing components **(personnel communication with Juan Valcarcel)**. The splicing failure seen in the first series mutagenic oriented FLEx-EGFP-Neo constructs could have resulted from the failure to observe such distance requirements. In this case, the distance impacted not only the recognition of splicing and polyadenylation signals but disrupted precise communication between the synt pA and synt sA signals is normally critical for recognition of the synthetic terminal exon in the FLEx cassette. Consistent with distance

being important, insertion of IRES-RFP sequence between the sA and pA in the cassette reversed this splicing failure. In this case, mutagenic oriented FLEx was able to capture the downstream region of the transcript and prevented EGFP expression.

We next tested if a the FleX cassette containing the IRES-RFP may possibly fail to mutagenize due to effects of synthetic exon length on splicing, as suggest by a previous study **(87).** The GFP assay results showed that this was not the case. Additionally, we considered that the increased distance between lox sites generated by IRES-RFP insertion could decrease the cassette inversion capacity **(27)**. However, Cre recombination efficiency test results in selected clones stably transfected with this cassette indicated that the cassette is competent in both orientations. We were also aware that intronic insertions could mainly result from alternative splicing, but we did not detect any alternative splicing. Nevertheless, it was difficult to determine the cassette orientation because of poor RFP expression following the electroporation. In contrast to the transient transfection results, poor, and/or nearly no expression of the RFP marker gene was observed in mouse E14TG2a cells transfected with the second series of mutagenic oriented FLEx-IRES-RFP-EGFP-Neo. This could be attributed to the low emission rate of the DsRed (RFP) marker gene that is used in the second cassette series and to slower fluorescence maturation compared to EGFP in mammalian cells. The advantage of using IRES for this kind of study is that translation of the reporter gene is independent of the endogenous initiation codon. RFP translates from the IRES start codon. It is also possible that IRES may be inefficient in this application as it has not been tested in ES cell studies prior to this project. This could explain the expression failure.

3.2 The presence of mammalian transcription termination signals provides high fidelity to the mutagenic oriented FLEx

Gene expression control through a recombinase mediated gene trapping cassette is not unique to this PhD project. Two other research groups generated gene trapping cassettes for conditional gene inactivation based on recombinase mediated unidirectional DNA inversion **(91, 128)**. Dr. Xin's research group **(128)** constructed a reporter cassette

composed of a strong and functional sA, the human bcl2 intron2/exon3, an EGFP reporter gene and the bovine growth hormone pA signal, all flanked by mutant and wild-type lox sites placed in a configuration to enhance Cre mediated cassette exchange. In the pCAGGs-revGI-nlsDsRED vector construct, this cassette was inserted into the first intron of β-actin downstream of the CAGGs promoter in an anti-sense configuration, neutral at the expense of splicing, polyadenylation and EGFP expression. DSRed sequences with the nuclear localization signal nls were placed downstream of the cassette in a sense configuration with the CAGGs promoter. In the pCAGGs-GI- nlsDsRED vector construct, the cassette was placed in a sense orientation that allows the it to actively participate in mRNA processing events and results in EGFP expression. When HEK293 cells were transiently transfected with the pCAGGs-revGI-nlsDdRed vector, only DsRed gene expression was observed 24 hours after transfection. In the anti-promoter oriented cassette, the RNA processing elements were neutral and not recognized. Thus the cassette had no effect on DsREd gene transcriptional activity. Whereas, in cells transiently transfected with pCAGGs-GI-nlsDsRED vector, the sense oriented cassette captured the DsRed transcript in the presence of functional sA and pA. Based on these experimental results, the gene trap cassette is reliable in both orientations. The antisense oriented cassette was totally neutral, it did not interfere with downstream gene transcription. On the other hand, the sense oriented cassette created a mutation at the RNA processing level that prevented downstream DsRed transcription.

Following Cre-mediated cassette inversion in both kinds of transfected cells, the results contradicted previous outcomes. After Cre plasmid co-transfection, the reverse oriented cassette turned and GFP expression was seen in the cells 24 hours after induction. At the same time, both RFP positive clones were seen along with yellow clones, both RFP and GFP positive **(128)**. In the case of sense oriented cassette inversion, transfected cells expressed GFP following Cre induction although it was previously shown that antisense oriented cassette transfected cells were only RFP positive. All these contradictions could result from the transient transfection method used to introduce the trapping cassettes and Cre plasmids. These plasmids may not have integrated into genome in a stable way and thus were temporarily expressed and lost. At the same time,

it would be very convenient to make stable cell lines with the same constructs and Cre plasmid using electroporation to avoid the limitation of transient transfection efficiency. Dr.Schnutgen's research group generated another kind of gene trapping cassette, pFlipROSAβ-geo, that totally distrupted the retinoblastoma gene through a first intronic insertion **(91)**. Thus the question is, which part the FLEx cassette, developed in our laboratory, distinguishes it from other cassettes used in the same gene trapping strategy?

In this PhD study, the gene trapping strategy was innovative for two major reasons: (1) high mutagenesis capacity at the level of transcription due to the utlization of mammalian transcription termination signals; and (2) amenability to Cre-mediated recombination. The first generation of FLEx cassettes did not prevent the production of downstream EGFP mRNA when inserted antisense into the first intron of the β-actin gene. Thus the cassette was reliable in terms of neutrality. The failure of the mutagenic oriented cassette only stemmed from the unfavorable arrangement of the splice acceptor and polyadenylation signal in the cassette. The distancing sA and pA in the cassette by adding IRES-RFP prevented this pitfall. Both ES cells both stable and transient transfected with mutagenic oriented FLEx cassettes confirmed this fact. Remarkably, neither transfection nor Cre efficiency seemed to interfere with the expected results in stable E14TG2a clones containing either orientation of the FLEx cassette at 24 hours after transient transfection of the Cre plasmid. Indeed, high frequency of cassette inversion seen 24 hours after lipofectamine transfection of the Cre plasmid showed that although the risks of unstable Cre integration and expression are still present, they did not affect this particular experiment.

This project was aimed at developing a versatile tool to for large-scale mutagenesis in the mouse. This objective requires that the cassette is able to control complex and highly coordinated transcriptional events in mammalian systems. In both yeast and mammalian systems, transcription termination, the last stage of RNA Pol II transcription that involves dissociation of both the newly synthesized pre-mRNA and RNA pol II enzyme from the template DNA, has a regulates gene expression as well as splicing and polyadenylation. Although a functional pA signal is necessary for effective transcription termination, RNA pol II does not terminate at the pA signal. *In vitro*

experimental systems coupling RNA pol II transcription and polyadenylation pointed out the pA signal only directs the slowing down of the RNA pol II enzyme. Therefore, we designed the FLEx cassette with the special signals that enhance transcription termination.

One remaining question to be answered in this project was, Are the transcription termination signals really necessary for the mutagenicity of the cassette? To determine this, FLEx gene trapping cassettes were engineered without these signals and assayed with GFP. The absence of the signals did not affect the neutral character of the cassette, as expected. However, in the sense orientated FLEx cassette, the absence of the signals positively affected downstream EGFP mRNA transcription. GFP expression is not normally seen in cases of mutagenesis by this cassette. This means that the cassette lost its mutagenic character in stable transfected E14TG2a cells, indicating that the termination signals are a critical part of the cassette.

In conclusion, efficient splicing and polyadenylation have key roles in controlling expression of the trapped gene. In particular, the termination signals downstream of the pA signal served as check point factors.

3.3 CoTC is a strong component of mammalian transcription termination

The usage of the transcription termination signals, CoTC and MAZ4, was the novel part of this PhD project. Neither of these signals had previously been applied in this way. Moreover, there had not been any scientific evidence showing the dominance of CoTC versus MAZ4 in transcription termination. Mammalian transcription termination, in additional to translation control, strictly coordinates the response to protein demand. Proper transcription termination is critical not only to preventive the production of truncated mRNA due to early stoppage of Pol II enzyme, but to regulate neighbouring gene expression by preventing a late Pol II stop. In mammals this process is less understood than other transcription stages. Further studies are required to untangle this complexity.

Two main models have emerged to describe the mechanism of transcription termination, the anti-terminator and the torpedo model. Recently, one group indicated that the actual mechanism is a combination of these two models. This is based the discovery that termination showed extensive heterogeneity in most mammalian genes, occurring at various positions along the pA signal instead of at a single site. This supports a requirement for heterogenic auxillary termination elements such as CoTC in the β globin gene or MAZ4 in the C2 complement gene downstream of the poly A signal for the control of gene expression **(36,107)**.

In this study, we used the different combinations of CoTC and MAZ4 to find out what works best to create a total mutation at the RNA transcription level. GFP assay results with the second generation of mutagenic FLEx cassettes indicate that successful EGFP trapping occurs in the presence of CoTC rather than MAZ4. MAZ4 operated at a rate that was 50 % less than CoTC. These observations were further supported theabsence of GFP expression in E14TG2a cells transfected with a mutagenic FLEx cassette containing the BSD marker gene and CoTC. It is possible, however, that this principle does not universally apply to all mammalian genes. In other ES cell types, and/or in other gene, MAZ4 may be useful for preventing Pol II sliding down the RNA. After presenting our lastest findings with the FLEX constructs containing BSD selection marker at the EUCOMM February 2006 meeting, Dr. Proudfoot's research group announced that the strong four G repetitive sequences, MAZ4 were not sufficient to promote transcriptional termination **(36)**. This result supported our results in mouse E14TG2a ES cells. They tested MAZ4 in the β-globin gene by substituting it for CoTC, 800 bp downstream of the pA signal and transfecting HeLa cells with this construct. They found that compared to CoTC, MAZ4 worked 57% less efficienctly in this gene. The β-globin gene is part of a closely spaced gene cluster. In this cluster, the intergenic elongation of the pol II enzyme is controlled by pausing signals in genes such as Gγ globin within the cluster on chromosome 11. The Aγ globin gene in the same cluster has an identical pausing mechanism, but due to its weak pA signal, it is weaker than that of the Gγ globin gene. Pausing activity is strongly dependent on a functional, strong pA signal **(81)**. Our

observations that the mutagenic FLEx cassette with the MAZ4 signal is weak, illustrate a failure of weak pA signal within the cassette to reliably pause or slow the pol II enzyme.

Proximity of the MAZ4 to strong pA signal was effective to complete transcription. This was expected as it has been shown that control of immunoglobulin gene termination in the presence of a pause site 1 kb downstream of the pA signal does not affect termination. The efficiency of MAZ 4 from 50 bp distance downstream of pA signal has also been tested and found to have no negative influence on pausing efficiency **(80)**. However, the possibility of negative effect at very close distances has not yet been investigated. Trapping of GFP in the second (IRES-RFP constructs) and third generations of FLEx cassettes (BSD constructs) that contain both signals may result from the close distance. Placement of MAZ4 immediately downstream of the pA signal did not result in downstream GFP transcription capturing. On the other, placement of MAZ4 downstream of both the pA did not interfere with EGFP transcription termination. This suggests that MAZ4 may be enhancing termination in combination with CoTC. Of course in these experiments, CoTC function was independent of both pA proximity and strength, and the results seen here may have been due to this.

This data raises the question of what pA signal types are utilized in the mouse genome. Investigations were done of pA signal variations among the mouse genes. Up to now, the most interesting and reliable pA signal mapping study was completed in 2005. Based on NCBI databases, including gene bank, Unigene, and dbEST, 16.282 pA sites were mapped in 11.255 mouse genes using the consensus hexamer AAUAAA or its closed variants and the heterogeneity of cleavage sites. This research concluded that gene function, structure, cell/tissue specificity, and the developmental stage of expression, direct preferences for pA signal type and usage **(103)**. To make this choice, specific cis acting elements and trans acting factors are critical. Another investigation of the pA signal preferences in human genes turned attention to pA efficiency and the factors determining this preference. Indicators of pA strength are DSE structure and the distance of this U or G rich sequence to the pA, AAUAAA hexamer or its variations **(54)**.

When we looked for pA signal usage preferences for the construction of an efficient gene trapping cassette, we found no published rules. For mouse mutagenesis, SV40 triple pA, bgh pA, and MMLV U3 pA are the most commonly used signals in gene trapping cassettes **(111)**. The common feature of all these signals is that their strength depends on the distal proximity of the DSE to the pA signal hexamer. In this PhD project, it was not possible to test every kind of pA signal, but we can now offer some rules for uninterrupted cassette functionality. For conditional first multipurpose allele generation with FLEx, the first rule about the pA element in the gene trapping cassette is that it should be unidirectional so that it does not interfere with endogenous gene transcription. In this project, usage of the strong pA signal, bidirectional SV40pA, in the third generation of neutral oriented FLEx cassettes was important to monitor the dangerous effects of SV40pA. The percentage ES cells that expressed GFP was 50 % decreased in the presence of SV40 compared to the synthetic pA signal. For large-scale gene trapping events, incorporation of a functional, but not necessarily strong, pA element in along with an efficient transcription termination signal, such as CoTC, will be practical.

3.4 Gene targeting – definite but not accurate for all routine mouse genome manipulations

Here, we tested the gene trapping FLEx cassette by integrating it into the β-actin intronic region of the CAGGs-EGFP-Neo construct**.** The best configuration of the FLEx cassette was determined by assaying GFP expression in FLEx cassette series-transfected E14TG2a ES cells. The FLEx cassette is now ready to use for targeted trapping. It has strong RNA transcript processing signals (synt sA and synt pA), a reliable terminator signal (CoTC), a suitable marker gene for mammalian systems (blasticidin), and is able to be utlized with Cre recombination. It is also reliable in both orientation.

To simplify this application, we selected the X-linked Jarid1c gene, which is single-copy in male E14TG2a ES cells from the 129 Ola strain. Our first goal was to generate a Jarid1c conditional allele via single step homologous recombination. Unfortunately, homologous recombination with the successfully engineered and

subcloned Jarid1c BAC targeting construct was not achieved in E14TG2a ES cells. The major limitation of this strategy was the high rate of random vector integration.

Gene targeting is dependent on sequence similarity (homology) between the exogenous DNA (targeting construct) and the endogenous DNA (targeted gene locus). Gene targeting is the preferred to strategy to specifically modify of mouse genome. However, it is still impractical for routine applications. This application is totally dependent on endogenous cellular DNA rearrangement machinery, such as the repair system. Compared to other eukaryotic organisms, like drosophila and yeast, the frequency of homologous recombination during foreign DNA integration into the targeted locus is extremely low versus illegitimate recombination, 1000-10.000 times less frequent. The versatile targeting ratio is infinite because the preferences of recombination type during foreign DNA integration into the mammalian genomes, as well as the factors directing these preferences, are still ambiguous. In addition to cellular factors, experimental factors influence targeting frequency and exhibit variability from gene to gene. Analyzing previous attempts to increase gene targeting helped to clarify the reasons behind the Jarid1c targeting difficulties. We focused on putative factors that impede targeting, like an impervious gene locus, improper targeting construct, poor ES cell conditions or inefficient construct delivery.

Chromatin status and Jarid1c expression levels were critical considerations during the planning stage of Jarid1c targeting in male ES cells. Affymetrix gene chip expression data from Dr.Sandra Lubitz **(personnel communication)** indicated that Jarid1c is expressed in E14.1 (129 strain ES) cells. In addition, the gene was trapped in a random fashion in C57BL/6Jx129S6/SvEvTac F1 derived ES cells. All these were male origin **(39,91)**. In this project, we used E14 (129 strain) ES cells. In contrast to the expected results of homologous recombination, we found that illegitimate recombination always occurred between the trapping exogenous cassette and the first intronic region of the gene. Therefore the possible negative effect of an imperveneous gene locus was not the reason for recombination failure in the previous study. Unfavourable distribution of heterology between the targeting vector and targeting locus may be the culprit **(23).** In this project, our Jarid1c BAC DNA source was from the C57BL/6J strain, the only

alternative strain that has been annotated in the Ensembl genome information. We also increased the length of homology arms on the targeting construct. This was not expected to cause problems in nuclear diffusion efficiency because a previous study indicated that 10-25 kb arm length has no effect **(23)**. We also determined the sequence divergence of both arm regions from C57BL/6J and two substrains of 129 DNA. The 30 kb Jarid1c transcript from C57BL/6J was aligned against 129/SvlmJ and 129/SvJ **(sequence information Daniel Rios from EBI).** Information from 129 substrains was not available for 7.3kb of theleft arm and 6.1kb of the right arm. In summary, we could not obtain information about the homology arm from the 129 DNA source. In this case, the left homology arm length could be 2.7 and right arm could be 3.9 kb. The probable high degree of heterogeneity on the homology arm could complicate Jarid1c targeting.

For mouse genome modifications, ES cells are always the preferred cell type. Activity and duration of the cell cycle is dependent on the cell state. ES cells are highly proliferative due to more than half of its cell cycle being the S phase. Gene trapping is dependent on the homologous machinery that is mostly active in the cell cycle phase G2/S. Therefore, a high rate of homologous recombination is found in ES cells. However, ES cells can lose this state. Sponatenous differentiation comes from high passage rates and potential from the enzymatic passaging method. Trypsin treatment may create such a handicap for gene targeting **(29)**. For the second targeting, we used more healthy, highly proliferative and less passaged IB10 cells. Unfortunately, this did not change the targeting fidelity of this gene.

Our last experimental consideration for the Jarid1c targeting case was the technical efficiency of ES cell transfection. DNA delivery method is another essential factor that influences the chance of foreign DNA integration into the host genome. The main parameters for high efficiency transfer of nucleic acids to specific mammalian cells are low toxicity, the ability to keep the nucleic acid stability, and the practicality of usage. Common DNA transfection methods can be classified as: chemical (DEAE dextran, calcium phosphate-DNA co-precipitation, liposome mediated DNA transfection); physical (electroporation, direct microinjection of DNA); and viral (infection). The disadvantages and advantages of each of these delivery methods is clear, but information

is limited on how they affect transfection efficiency and DNA stability. There is no information about how they influence nuclear events, e.g., recombination following exogenous DNA transfer.

In this project, we used electroporation to transfer DNA into the cytoplasm by emitting electric pulses that affect cell permeability. Its efficiency has been proven in mammalian cells **(31)**. However, whether electroporation can affect DNA stability in cytoplasma and its transfusion into the nucleus is a key question in gene targeting studies. In 1980, when mammalian gene targeting studies first started, one study indicated that DNA directly microinjected into the host cells stably integrates in only 20% of the cells. Electroporation is more advantageous than this because it results in a larger number of transfected cells **(10,31).** Today, transfection rate has been improved thanks to advanced nucleic acid delivery techniques. However, the problem still remains that these techniques introduce exogenous DNA into the cytoplasm but they can not affect nuclear transfusion. One research paper showed that foreign DNA is rapidly degraded upon transfection **(110)**. In nucleus, it is not clear if it undergoes modifications such as large deletions and point mutations, before or after integration into the genome **(110)**. It was also shown that foreign DNA in the nucleus can be efficiently circularized, but only in the case of microinjected DNA **(110)**. Unfortunately, nuclear events cannot be controlled by the today's transfection methods. These uncontrollable events might have played a role in our Jarid1c targeting difficulties.

Another potential reason for this targeting failure could be the presence of illegitimate recombination hot spots on the mammalian chromosomes. Major groups are now working on the molecular mechanism of the illegitimate recombination and the tendency of the specific DNA regions to promote these kind of genetic arrangements **(10,91)**. It was possible to trap the intronic region of Jarid1c between coordinates 147.574.367-147.574.426 on the X chromosome using the intron trapping cassette, rFlpROSABetaGeo (91), but it was impossible to target the intronic region of Jarid1c between coordinates 147.577.072-147.577.167 using the FLEx-Jarid1c targeting construct. This difference may be an example of a hot spot effect.

In summary, our work has produced a gene trapping cassette that shows high fidelity in GFP trapping. We also have demonstrated that transcriptional events and elements enhance mutagenesis capabilities of the FLEx gene trapping cassette. This was a novel finding for gene trapping application. In the future, it will be important to direct homologous recombination in ES cells by exerting control over nuclear events. Dr. Dominguez-Bendala's effort to increase targeting frequency in ES cells through transfection of the Rad51 plasmid was the fascinating example for this theory **(24)**.

4. METHODS

4.1 DNA protocols

4.1.1 Small-scale preparation of plasmid DNA

This protocol has 3 basic steps; (1) alkaline lysis of bacterial cells with SDS, (2) plasmid DNA isolation, and (3) washing and purification of the plasmid DNA. From a 1 ml culture, the typical yield for high-copy plasmids (pUC, pBluescript vectors) is up to 20 μg, for mid-copy (pACYC and derivatives) and low-copy plasmids (pSC101 vectors) it is 10-15 μg. Bacterial culture volume can be 2 ml for low-copy plasmids. Further information for other critical considerations such as construct size can be gathered from Gene Bridges Company.

△ Inoculate a single colony into 1 ml LB with suitable selection agents, antibiotics, and incubate at 37^0C with vigorous shaking (~1000 rpm) 12-16 hrs.

CRITICAL: If the selection agent is blasticidin, please use low-salt LB with pH 8.0.

△ Spin the overnight culture 1 min at 13,200 rpm at RT, remove the supernatant and resuspend the bacterial pellet in 200 μl buffer P1.

△ Add 200 μl buffer P2 and mix by inverting, at this point 5 min RT incubation can be done for proceeding the lysis reaction but no more than 5 min.

△ Add 200 μl buffer P3 and mix by inverting. It can be left on ice for 20 min.

△ Spin 20 min at 13,200 rpm at rt, transfer the supernatant into an eppendorf tube containing 600 μl isopropanol. Vigorously mix and spin at max speed, 13,200 rpm for 20 min at RT. This precipitates the plasmid DNA

△Wash the pellet 75% ethanol, spin 1 min at 13,200 rpm at RT. Then remove the alcohol.

△ Dry pellet at 5 min 45^0C, then redissolve the DNA in a suitable volume of buffer, 20-30μl (TE, pH 8.0 for long term storage, 10mM Tris-Cl, pH 8.5 or sterile water).

4.1.2 Large-scale preparation of plasmid DNA

This procedure is based on a modified alkaline lysis of bacterial cells followed by binding plasmid DNA to an anion-exchange resin under low salt and pH conditions. RNA, proteins, dyes, and low-molecular-weight impurities are removed by a medium-salt wash. Plasmid DNA is eluted in a high salt buffer and then concentrated and desalted by isopropanol precipitation. Plasmids up to approximately 150 kb can be purified with this method, but plasmids larger than 40-50 kb may show reduced elution efficiencies. From 100 ml culture the typical yield for high-copy (pUC and pBluescript vectors) and mid-copy plasmids (pACYC and derivatives) is 300-500 μg. The expected yield from 500 ml of low-copy plasmids (pSC101 vectors) is100-500 μg. Further information can be gathered from the Qiagen plasmid purification handbook.

△ Inoculate a single colony into 100 ml LB for high-copy, or 500 ml LB for low copy with suitable selection agents, antibiotics, and incubate at 37^0C with vigorous shaking (~300 RPM) 12-16 hrs.

CRITICAL: If the selection agent is blasticidin, please use low-salt LB with pH 8.0.

△ Harvest the bacterial cells by centrifugation at 6000 rpm for 15 min at 4^0C

(Beckman JLA 16.250 rotor)

△ Resuspend the pellet in 10 ml buffer P1, vortex vigorously

△ Add 10 ml buffer P2, mix gently through inverting and incubate at rt for 5 min to complete lysis reaction.

△ Add 10 ml pre-chilled buffer P3 and mix gently through inverting.

△ Pour the lysate into the barrel of the filter cartridge, incubate at rt for 10 min.

△ Gently insert the plunger into the cartridge and filter lysate into the 50 ml falcon tube.

CRICTICAL: If the DNA will be used for cell transfection or the other approaches similar sensitivity, endonuclease buffer (2.5 ml) has to be added into the filtered lysate, and incubate 30 min on ice. Then the rest of protocol steps should be followed.

△ Equilbrate the QIA-tip 500 by applying 10 ml QBT buffer and allow the column to empty by gravity flow.

△ Apply the lysate to the tip 500 and allow it to enter the resin by gravity flow.

△ Wash the tip 500 with 2x 30 ml wash buffer, QC.

△ Elute DNA with 15 ml elution buffer, QN.

△ Precipitate DNA by adding 10.5 ml (0.7 volumes) isopropanol to the eluted DNA. Mix, and not necessary to immediate centrifuge. Incubation at rt increases the yield.

△ Centrifuge at 7000 rpm, 1 hr, at 4^0C

(Beckman JS 13.1 rotor and 30 ml Corex glass tube with adaptor)

△ Wash the pellet with 5 ml %75 ethanol or 5 ml 70% endotoxin-free ethanol supplied by kit.

△ Air dry the pellet for 10 min, then then redissolve the DNA in a suitable volume of buffer, ~250-300 μl (TE, pH 8.0 for long term storage, 10mM Tris-Cl, pH 8.5 or sterile water).

4.1.3 Mini-scale preparation of BAC DNA

This protocol is designed for BAC DNA purification from 2 ml culture. Further information can be gathered from the Epicentre BACMAX plasmid purification handbook.

△ Inoculate the single colony into the 2 ml LB with suitable selection agents, antibiotics and incubate at 37^0C with vigorous shaking (~1000 rpm) 12-16 hrs.

CRITICAL: If the selection agent is blasticidin, please use low-salt LB with pH8.0.

△ Spin the overnight culture 1 min at 13,200 rpm at RT, discard the supernatant

△ Add 200 μl chilled BACMAX solution 1 to the pellet. Vortex vigorously to completely resuspend the pellet.

△ Add 400 μl BACMAX solution 2 to suspension, mix by inverting the tube 2-3 times very gently and incubate no more than 5 min at rt to complete lysis reaction.

△ Add 300 μl chilled BACMAX solution 3, mix by inverting the tube 2-3 times very gently. Incubate on ice for 15 min.

△ Centrifuge at max speed, 13,200 rpm for 15 min at 4^0C, transfer the supernatant in an eppendorf tube

△ Centrifuge at max speed, 13,200 rpm for 15 min at 4^0C, transfer the supernatant in an eppendorf tube, add 540 μl (0.6 volumes) room temperature isopropanol, mix thoroughly by inverting the tube 4-6 times.

△ Centrifuge at max speed, 13,200 rpm for 15 min at 4^0C, carefully remove the isopropanol, dry the pellet at rt for 5 min.

△ Redissolve the DNA completely in 250 μl TE buffer.

△ Add 1 μl Riboshedder RNAse blend to the tube and incubate at 37^0C for 30 min.

△ Add 250 μl chilled BACMAX solution 4, mix thoroughly by tapping the tube and incubate on ice for 15 min.

△ Centrifuge at 13,200 rpm for 15 min at 4^0C.Transfer the supernatant to an eppendorf tube, add 1 ml of absolute ethanol to the recovered supernatant, mix gently by inverting the tube 4-6 times.

△ Precipitate the DNA by centrifugation at 13,200 rpm, for 15 min at 4^0C.

△ Air-dry the pellet at room temperature for 5 min, and then redissolve the DNA in 25 μl (TE, pH 8.0, or 10 mM Tris-Cl or sterile water)

4.1.4 Preparation of mouse genomic DNA from mouse E14TG2a cells from 6 well-plate

● Aspirate the media from 6 well plates containing confluent ES cells, wash once with 3ml PBS, aspirate.

● Add 800 μl PBS, and scrape by a soft movement with sterile scraper. Take the cells with a micropipette and put into an eppendorf tube.

● Alternative to the scraping, put 300 μl trypsin, incubate 1-2 min at 37^0C, then add 300 μl medium to inactivate the trypsin. Put into an eppendorf.

● Centrifuge at 4000 x g, at rt for 4 min

● Wash with 1 ml PBS by spin at 4000 x g at 4^0C for 4 min

● Remove PBS, put 1ml fresh PBS. Allocate 250 μl for genomic DNA isolation and 750 μl for freezing to further aims.

- Add 500 μl lysis buffer and incubate overnight at 37^0C

- Next day, put 500 μl phenol/chloroform/isoamylalcohol (25:24:1) and rotate 1 hour on a rotating drum.

- Centrifuge at 10,000 x g at rt for 10 min, take the aqueous phase containing DNA carefully not to contaminate with lower phase protein, phenol.

- Add 1 volume chlorophorm (1:1) approximately 400- 500 μl and rotate 1 hour on a rotating drum.

- Centrifuge at 10,000 x g at rt for 10 min, take the aqueous phase containing DNA carefully not to contaminate with lower phase protein, phenol, chlorofom and other impurities.

- Add 25 μl 3M potassium acetate, and 400 μl isopropanol, mix. Incubate for 5 min at -80^0C, centrifuge at 10,000 x g for 15 min at 4^0C.

- Wash the pellet with 75% ethanol, spin at 10,000 x g for 5 min at 4^0C.

- Remove the ethanol, and dry the pellet for 5 min at rt. Add 100 μl TE buffer, pH 8 and incubate in a water bath for few hours at 37 ^{0}C, then overnight at rt.

4.1.5 Preparation of mouse genomic DNA from mouse E14TG2a cells from 48 well-plate

For routine analysis, it seems that the 6-well-plate approach is not practical, also the DNA concentration from 96 wells is not enough. Therefore, I modified the 48-well-plate protocol from Yale University, Animal Genomics services. With this approach at least two rounds southern blot analysis is possible.

- Aspirate the media from 48 well plates containing confluent ES cells, wash once with 500 μl PBS, aspirate.

- Add 150 μl lysis buffer, seal plate with parafilm, put in a box containing wet towel and incubate at 55^0C for overnight.

- Next day, transfer the lysate to an eppendorf tube, add 75 μl saturated NaCl and shake vigorously.

- Spin at 4000 x g for 15 minutes; transfer the supernatant to a new tube.
- Spool out the DNA with a tip but it was cut not to shear the DNA. Wash the pellet with 75% ethanol, centrifuge 10,000 x g for 5 min at 4^0C.
- Resuspend in 50 μl TE buffer, pH 8.

4.1.6 Restriction Enzyme Digestion

Restriction enzymes specifically cut plasmid and genomic DNAs into fragments for manipulation, analysis and further assays. Although each restriction enzyme has its own optimum conditions, some rules for the appropriate DNA digestion are universal. The most critical rule is that the DNA should be purified. As a rule, 1 unit of restriction enzyme cuts 1μg DNA in 1 hour. Actually, for plasmid DNA digestion, the enzyme unit should be two times (1μg DNA/2 unit enzyme), and for genomic DNA digestion, the enzyme unit should be 5 times higher than the DNA concentration (1μg DNA/ 5 unit enzyme). Incubation time and temperature is dependent on factors like enzyme type and the DNA nature. Further information about digestion conditions can be gathered from NEB, Stratagene, Andras Nagy laboratory, and Cold Spring Harbor protocol pages. Also critical points for suitable enzyme selection for genomic DNA can be read from Nagy pages.

Analytical reaction	**Preparative reaction (eg: fragment isolation, southern blot)**
DNA x μl(0.2-0.5μg)	**DNA** x μl(10-20 μg)
10 X Buffer 1μl	**5 X Buffer** 5 μl
R.E x μl **(**x unit)	**R.E** x μl **(**x unit)
If necessary 0.1 X BSA	**If necessary 0.5 X BSA**
Add to 10 μl sterile water	Add to 50 μl sterile water

4.1.7 Agarose gel electrophoresis

Gel electrophoresis is a method to separate DNA, RNA or protein molecules by size. This is achieved by moving negatively charged nucleic acids through an agarose

matrix with an electric field at rates that are inversely proportional to the size, and conformation (linear or circular) of the molecules. Shorter molecules move faster than longer ones. From the slowest to the fastest, plasmids in different conformations migrate at different rates; nicked open circular DNA, linear DNA, relaxed circular DNA, supercoiled and the fastest, denaturated supercoiled DNA.

Gel electrophoresis can be used for the separation of DNA fragments of 50 bp to several megabases. Small nucleic acids are better separated by polyacrylamide gel electrophoresis. Larger nucleic acids are better separated by pulse-field electrophoresis. For appropriate resolution, the agarose gel concentrations are summarized below.

Linear fragment size	**% agarose**
30 kb to 15 kb	0.4%
15 kb to 1 kb	1%
2.5 kb to 0.5 kb	2%
< 100 bp	3%

The most common electrophoresis buffers are Tris/acetate EDTA(TAE), the best for resolution of larger fragments, and Tris/Borate EDTA(TBE). As for the loading buffers, bromophenol blue is commonly used. 1x bromophenol blue is generally added to the DNA. For the BAC DNA electrophoresis, orange G is useful. The running voltages are 75 volts for plasmid DNAs and 50 volts for BAC DNA. The rule is 5-10 volts/cm. DNA. The DNA is visualized by 20 ng/ml ethidium bromide.

4.1.8 Phenol-chloroform extraction of plasmid DNAs

This technique is used for purifying DNA from proteins. Minimum start DNA volume is 100 μl, a lower volume should be diluted with TE buffer or water up to 100 μl. For the first step phenol:chloroform:isoamylalcohol(25:24:1) is added in a 1:1 volume to the DNA and mixed, small fragments can be mixed by vortex. Centrifuge for 5 min, at 16,000 x g, at RT. The aqueous phase is removed, then 1 volume chloroform is added, mixed and centrifuged under the same conditions. Then DNA precipitation is performed. For this the final concentration of 0.3 M sodium acetate is added, mixed, and absolute

ethanol is added. Following the 5-30 min -80^{0}C precipitation, the mix is centrifuged at 15 min, 16,000 x g. The pellet is dried at 45^{0}C for 5 min or at RT for 10 min, and redissolved in a suitable volume of TE, pH 8.0, 10 mM Tris-Cl or sterile water.

4.1.9 Ligation

DNA ligations are performed by incubating vector and insert DNA (DNA fragment from gel or PCR product) appropriately 1 (vector): 2 (insert) in the presence of T4 ligase buffer, enzyme and sterile water. The reaction volume is 10 μl. In this project either blunt end or sticky end ligation, 4^{0}C reaction temperature (refrigerator) and 24 hour incubation time were used.

Ligation reaction

Vector DNA x μl

Insert DNA 2x μl

10 X T4 DNA ligase Buffer 1.07 μl

T4 DNA ligase 0.5 μl **(**5 unit)

Add to x μl sterile water to set up 10 μl

4.1.10 TOPO TA Cloning

TOPO TA cloning provides a highly efficient one-step cloning strategy for the direct insertion of Taq polymerase amplified PCR products into a plasmid vector. No ligase, no post-PCR procedures (purification and precipitation) and no PCR primers containing specific sequences are required.

Cloning reaction

Vector DNA 1 μl

PCR product 4 μl

Salt solution 1

Add to 5 μl sterile water to set up 6 μl

5 min-overnight incubation (30 min is generally enough) for cloning reaction. The salt buffer should be diluted 4 fold to prepare 300 mM NaCl, 150 mM $MgCl_2$ if electrocompetent *E.coli* are used. 2 μl reaction sample is electrophorated or transformed into chemical competent *E.coli*.

4.1.11 Polymerase chain reaction (PCR)

This technique permits us eto xponentially amplifying specific DNA regions through enzymatic replication *in vitro*. PCR is carried out in 3 main steps; (1) denaturation of the template DNA, (2) primer annealing, and (3) extension/elongation. It consists of 20 to 35 cycles.

In this PhD project, PCR was performed on a Stratagene Robocycler machine (2003) or an Eppendorf PCR machine (2005-2006).

PCR mix (50 μl rxn)

DNA template x μl (10 ng for plasmid DNA, 100 ng for genomics DNA)

dNTP 1 μl (10 mM)

F primer 1 μl (10 pmol)

R primer 1 μl (10 pmol)

5X Taq DNA polymerase buffer 5 μl

Taq DNA polymerase 0.5 μl (2.5 unit) and X μl water to complete 50 μl.

Denaturation was always carried out at 95^0C. 30 cycles were performed and annealing temperature was optimized based on the primer Tm for each reaction. The last cycle was always 10 min at 72^0C.

4.1.12 PCR product purification

For the molecular cloning reactions, PCR products were purified using the agarose gel extraction method to obtain 70 bp- 10 kb DNA from standard of low-melt

agarose gel. For Red-ET cloning, PCR products were precipitated (section x DNA precipitation).

- Cut the DNA fragment from agarose gel with a sterile scrapel
- Weigh the gel slice, as a rule, add 3 volume of buffer onto the 1 volume of gel (1 volume=100 mg ~ 100μl).
- Dissolve the gel completely by incubation at 50 for 10 min, every 2 min mix by vortex.
- Add 1 volume of isopropanol to the dissolved gel. This step increases the yield of DNA fragments < 500 bp and > 4 kb.
- Apply the sample onto the QIA quick spin column, and centrifuge for 1 min at rt, at 13,000 RPM. In order to remove the traces of agarose, 500μl buffer QG can be added on to the column and centrifuge done again for one min at rt, at 13,000 RPM.
- 750 μl buffer PE is added onto the column after 1 min centrifugation at rt at 13,000 rpm, additional centrifuge repeated at the same condition to remove the alcohol completely.
- 30-50μl buffer EB (10 mM Tris-Cl, pH=8.5) or sterile water is added to the center of the column to elute the DNA. After 1 min rt incubation, centrifuge for 1 min at rt, at 13,000 RPM.

4.1.13 Red-ET cloning technology

Red-ET cloning technology relies on homologous recombination *in vivo* in *E.coli*. Homologous recombination occurs through homology regions, namely stretches of DNA shared by the two molecules that recombine. The mechanism and the practical steps of this technology is detailed in chapter 1, section 1.2.3.2.1. In this thesis, this technology was used to insert a NheI restriction site to the mutagenic oriented pFLEx-IRES-RFP-GFP constructs which show different patterns of RNA pol II termination signals. Also, It was applied to create the Jarid1c gene targeting cassette.

4.1.13.1 PCR reaction

PCR mix (50 μl rxn)

DNA template x μl (1-5 ng)

dNTP 1 μl (10 mM)

F primer 1 μl (10 pmol)

R primer 1 μl (10 pmol)

5X Taq DNA polymerase buffer 5 μl

Taq DNA polymerase 0.5 μl (2.5 unit) and X μl water to complete 50 μl.

Denaturation was always carried out at 95^0C. 30 cycles were performed and annealing temperature was optimized based on primer Tm for each reaction. The last cycle was always 10 min at 72^0C. After checking 3 μl PCR product onto a gel, PCR products were pooled. It was not necessary to purify the product unless the template was replicative plasmid. For this kind of template, PCR products were digested with DpnI. For this digestion reaction, precipitate the 3 tubes x 50 μl PCR products, redissolve them in 80 μl, add 40 Unit DpnI and digest overnight at 37. Next day, through the ethanol precipitation clean the PCR product and resuspend this in 8μl sterile water. For transformation of PCR product, 3-4 μl PCR product is used (as a rule >0.2μg).

4.1.13.2 Transformation of the Red-ET expression plasmid (pSC101-BAD-gbaA)

- Pick at least 8 colonies carrying the target plasmid DNA or BAC. Inoculate the single colony into the 2 ml LB with suitable selection agents, antibiotics, and incubate at 37^0C with vigorous shaking (~1000 rpm) 12-16 hrs.
- Next day, set-up eppendorf tubes containing 1 ml LB medium with suitable selection agents, antibiotics. Inoculate with 40 μl of overnight culture. Incubate 2 hours, at 37^0C at 1100 rpm.
- After 2 hours incubation, make electro-competent cells. Centrifuge for 30 seconds at 11,000 rpm at 2^0C. Remove the supernatant by quickly tipping out twice and put on ice.

Add 1 ml ice-cold water, resuspend the pellet, centrifuge for 30 seconds at 11,000 rpm at 2^0C. Repeat in a quick manner this washing, and resuspending steps twice. At last wash, keep 30 μl supernatant, and resuspend the pellet and put on ice.

- 1 μl (0.3 μg/μl) Red/ET recombination protein expression plasmid (pSC101-BAD-gbaA) is added to this 30 μl bacteria suspension. Then, this is transferred to the chilled electroporation cuvvette.
- Electroporate at 1350 volt, 10 μF, 600 Ohms.(Eppendorf electroporator 2510,1 mm electroporation cuvvette)
- Resuspend the electroporated cells in 1 ml LB without selective agent, antibiotics. Incubate at 30^0C, for 70 min at 1000 rpm.
- After 70 min, plate 100 μl cells on LB agare plates containing the appropriate antibiotics for recipient bacteria and tetracycline (3μg/μl) for pSC101-BAD-gbaA). Protect the plates from the light, because tetracycline is negatively affected, and incubate overnight 30^0C otherwise pSC101-BAD-gbaA will be lost at 37^0C. For blasticidin reagent, the agar plates should be prepared with low salt LB, pH 8.5.

4.1.13.3 Transformation of PCR product

- Pick at least 8 clones from the pSC101-BAD-gbaA transformed recipient bacteria plates. Inoculate in 1 ml LB (or low-salt LB) with suitable antibiotics and tetracycline. Incubate at 30^0C at 1000 rpm for overnight.
- Next day, set-up eppendorf tubes containing 1 ml LB medium with suitable selection agents, antibiotics. Inoculate with 40 μl of overnight culture. Incubate 2 hrs, at 30^0C at 1100 rpm.
- Add half of the tubes 20 μl L-arabinose (10%) giving a final concentration of 0.1-0.2%. This will induce the expression of recombinase. Other half of tubes are used as control with no induction. All the tubes incubate at 37^0C at 1100 rpm for 45 min.
- After incubation, prepare the cells for electroporation.
- Electroporate 3 μl PCR product at 1350 volt, 10 μF, 600 Ohms. (Eppendorf electroporator 2510,1 mm electroporation cuvvette)

- Resuspend the electroporated cells in 1 ml LB without selective agent, antibiotics. Incubate at 37^0C, for 1 hour 10 min at 1100 rpm. Then plate 100 μl cells on LB agar plates containing the appropriate antibiotics carried by recipient bacteria and the PCR product.

CRITICAL: Electrocompetent bacterial cells for the best recombination results should be used immediately. On the other hands, they can be kept at -80^0C following the snap freeze in liquid nitrogen.

4.1.14 Southern Blotting

The basic idea behind southern blotting is to separate digested DNA electrophoretically then transfer the DNA pieces onto a positively charged nitrocellulose or nylon membrane.

4.1.14.1 Gel electrophoresis and gel treatments for neutral transfer protocol

Following restriction digestion of genomic DNA(4.1.5 preparative reactions), overnight electrophoresis is preformed. DNA samples are separated on an 0.7 % agarose gel prepared with Tris/Borate EDTA (TBE) buffer by applying 40 volt for 24 hours. Next day, the gel is treated with an EtBr solution (10 μl of 0.1μg/μl stock solution added to 1 ml sterile water) to visualize the DNA samples in the gel with UV light and a photograph is taken.

CRITICAL: Minimize the exposure of gel to UV light to avoid excessive nicking of DNA.

The next step is now the 3 steps gel treatment for transfer the DNA to the membrane.

- **Depurination:**

Place the gel into the 0.125 M HCl, this should cover the gel completely. Agitate gently for 15 min, or till the loading dye is turned to yellow (around 10 min). Rinse the gel with sterile water.

CRITICAL: Depurination is not required for DNA fragments < 10 kb.

- **Denaturation:**

Place the gel into denaturation buffer; this should cover the gel completely. Agitate gently for 15 min, and then put fresh denaturation buffer and keep going with another 15 min agitation. Rinse the gel with sterile water.

- **Neutralization:**

Place the gel into neutralization buffer; this should cover the gel completely. Agitate gently for 15 min, and then put fresh neutralization buffer and keep going with another 15 min agitation. Rinse the gel with water.

4.1.14.2 Blotting

Put the gel in 20xSSC for 10 min with gentle shaking. Label nitrocellulose membrane, or nylone. Nylone membranes don't need pre-wetting before use in blotting because they are hydrophilic. For large blots or for nitrocellulose membrane, first wet in water and then equilibrate in an SSC (the same gel treatment buffer). Arrange the suitable apparatus and allow DNA transfer to proceed overnight. Next day, take the membrane and mark slots with pencil on the membrane, rinse it with 6XSSC and vacuum dry for 2 hours at 80^0C.

4.1.14.3 Radioactive DNA probe

△ Denature (100 ng DNA and sterile water to 11μl) DNA solution at 99^0C for 10 min

△ Put on ice then add 4 μl high prime (enzyme, primer,nucleotide mix,Roche)

△ Add 5 μl (50μC) alpha ^{32}**P** dCTP, incubate for 1 hour at 37^0C

△ Centrifuge the G50 sephadex column for 2 min at 2500 g, at rt.

△ Apply the labelled DNA onto the column and centrifuge for 4 min at 2500 g, at rt.

△ Collect the flow through and determine the volume

△ Measure the radioactivity of 1 μl sample and calculate the total activity.

△ Use around $2x10^7$ CpM for labelling the 10x10 cm DNA blotted membrane.

4.1.14.4 Hybridization in tubes

△ Pre-wet the membrane first with water and then 25mM $NaHPO_4$.Roll and put into hybridization tube.

△ Add 8 ml hybridization buffer and pre-hybridize for 30 min at 62-65^{0}C.

△ Denature labelled DNA probe for 10 min at 99^{0}C after 30 min pre-hybridization step.

△ Remove the hybridization buffer, add 10 ml fresh hybridization buffer at suitable hybridization temperature with the membrane. Add labelled DNA probe and hybridize overnight.

△ Next day, pre-warm the buffer to 55^{0}C. Then rinse the membrane briefly. Then twice wash for 5 minutes at rt. Then, put wash buffer and incubate the membranes through rotating at 65^{0}C for 15 min, and then remove the wash buffer. Add fresh wash buffer, and repeat the incubation through rotating rotating at 65^{0}C for 15 min. Remove the buffer and check the membrane with the Geiger counter.

△ Remove the membrane, drain and wrap in saran wrap and expose to X-ray film (KODAK MR) for 72 hours in a film cassette at -80^{0}C.

4.2 RNA protocols

4.2.1 RNA preparation from mammalian cells

Although, some applications are tolerant of partially degraded RNA, extraction of intact and pure RNA which is not contaminated with proteins, DNA, organic solvents and other impurities is critical for downstream applications such as RT-PCR, Northern blotting, and RNAse protection assays. Therefore all items such as glasses and reagents should be RNAse free. Glass items can be treated with DEPC water for 24 hours then autoclaved. DEPC treated water should be prepared (0.1% solution, 1ml DEPC is added to the 1L water) and after 24 hrs they should also be autoclaved. In this PhD project, total RNA was isolated from mouse E14TG2a cells when they were confluent in a single well

of a 6-well-plate, and RNA was used for RT-PCR analysis of EGFP gene expression in the first series of FLEx-EGFP-neo random stables(E14TG2a cells).

- Grow ES cells in 6-well-plate to confluence
- Aspirate the media from 6 well plates containing confluent ES cells, wash once with 3ml PBS, aspirate.
- Add 800 μl PBS, and scrape by a soft movement with sterile scraper. Take the cells with a micropipette and put into an eppendorf tube.
- Alternative to the scraping, put 300 μl trypsin, incubate 1-2 min at 37^0C, then add 300 μl medium to inactivate the trypsin. Put into an eppendorf.
- Centrifuge at 4000 x g, at rt for 4 min.
- Wash with 1 ml PBS by spin at 4000 x g at 4^0C for 4 min.
- Add 300μl Trizol reagent onto the cell pellet to lyse the cell, pass the lysate several times through a pipette to form homogeneous lysate.
- Incubate the homogenized samples for 5 minutes at rt.
- Add 30 μl of chloroform per 300 μl of Trizol and shake the eppendorfs vigorously by hand for 15 seconds and incubate them on ice for 5 min.
- Centrifuge the samples for 15 min, at 4^0C, at 13,200 rpm.
- Take upper clear aquaeous phase to a fresh eppendorf and add 150 to 200 μl isopropanol (volume should be 1:1)
- Precipitate RNA via mixing and incubate them at rt for 10 min
- Centrifuge for 15 min, at at 4^0C, at 13,200 rpm
- Wash the pellet two times with 80% ethanol (one drop, quick spin, repeat twice)
- Dry the pellet completely and dissolve in suitable amount DEPC water through 30 min rt incubation. RNA samples should be stored at -80^0C.

4.2.2 Reverse transcription of RNA (cDNA synthesis)

Minimum and maximum template concentrations should be between 1ng-1μg for total RNA and 10 pg- 1μg for mRNA.

- Set up the following components in an eppendorf tube

1.2 μg total RNA(used in this PhD project) +2 μl(1.8 mg/ml) oligo(dT) + sterile water to 13μl

- Incubate 5 min at 70^0C, then put on ice for 5 min
- Set up the following mixture and add per eppendorf

dNTP 5 μl (10 mM)

RNAse out 1 μl

5X M-MLV reverse transcriptase buffer 5 μl

M-MLV reverse transcriptase 1 μl (more than 200 units)

- Mix gently, spin down, and incubate 1 hr at 42^0C
- Stop reaction at 70^0C for 15 min
- Add 1 μl RNAse H and incubate at 37^0C for 20min

For the PCR reaction, use 5 μl cDNA.

4.3 Protein protocols

4.3.1 Protein extraction from mammalian cells

- Grow ES cells in 6-well-plate to confluence
- Aspirate the media from 6 well plates containing confluent ES cells, wash once with 3ml PBS, aspirate.
- Add 800 μl chilled PBS, and scrape by a soft movement with sterile scraper. Take the cells with a micropipette and put into an eppendorf tube.
- Centrifuge at 1000 rpm at 4^0C for 5 min, remove supernatant
- On each cell pellet, add 50 μl protein extraction buffer

- Shock-freeze the cells in liquid nitrogen, and then put into the warm water(melt the cells). Additional two times repeat this freezing-melting steps
- Then centrifuge at 13,200 rpm at 4^0C for 15 min, take the supernatant.
- For the measurement of protein concentration using absorbance first 230 then 260nm, mix 5 μl protein extract and 0.1% SDS, use PMSF as a control.

Concentration(μg/ml)= 100 x(183xOD 230-75.8xOD260)

4.3.2 Polyacrylamide gel electrophoresis (PAGE)

5%Stacking gel (1 ml)	**10%Running gel (5 ml)**
1M Tris pH 6.8 (0.13 ml)	1.5M Tris pH 8.8 (1.25 ml)
30% AA (0.17 ml)	30% AA (1.7 ml)
10% SDS (0.01 ml)	10% SDS (0.05 ml)
10% APS (0.01 ml)	10% APS (0.05 ml)
TEMED (0.001 ml)	TEMED (0.002 ml)
sterile water (0.68 ml)	sterile water (1.9 ml)

- Prepare the stacking and running gel
- Prepare 3μg protein extract mixed with coctail buffer(protein extraction buffer) to 20 μl total volume. Add 1x protein loading buffer(4x stock).
- Incubate at 70^0C for 5 min before loading
- Load the samples and run gel at 70 volt for 1 to 3 hrs depending on the size of protein should be detected.

4.3.3 Western Blotting

- Blot the proteins onto the Watmann membrane for 45 min, 15 V in a semi-dry blotter
- Incubate the membrane in blocking buffer for 45 min at rt with gently agitating

- Add primary antibody (Roche GFP mouse antibody) 1:5000 into the blocking buffer (2μl Ab+10 ml blocking buffer), incubate overnight at 4^0C with gently agitating.
- Next day, wash 3 times 20 min (total 1 hr) with PBS/0.1% Twin20
- Then add secondary antibody (anti mouse antibody) 1:5000 into the blocking buffer, incubate 1 hr at rt with gently agitating.
- Follow another 3 times washing step
- For the enhanced chemiluminesence (ECL), mix 1:1 ECL solutions, luminal enhancer solution + stable peroxide buffer, put your membrane on it, cover and incubate for 5 min, drain on a piece of tissue, image the specific protein with LAS-3000 image system (FUJIFILM)

4.4 Cell Culture Protocols

4.4.1 Starting out with ES cells

General steps for handling ES cells are growth, maintenance, passing, freezing, and thawing. They should be conducted in a careful way to keep the ES cells in a pluripotent state. The quality of the serum, with addition of the LIF, is critical for keeping the cells healthy and undifferentiated. 37^0C and 5% CO_2 are suitable incubation conditions. Daily refeed is another critical process for keeping them healthy. As other molecular biology techniques, cell culture techniques require sterility and carefulness.

4.4.1.1 Thawing the ES cells

- Take the cell from freezers at, -80 for overnight, - 150^0C for short-term storage (1-2 months) or from liquid nitrogen tank for long-term storage. Quick thaw at 37^0C water bath.
- Transfer the cells to the 15 ml falcon tube, add 10 ml ES cell media, gently mix, make cell suspension and centrifuge at 1000rpm at rt for 5 min.
- Aspirate off the supernatant, mix again cells, make cell suspension.

- Add 10 ml ES cell media, through pipetting make them single cell, and plate out them in a 10 cm cell culture dish. $1x10^6$ cell should be seeded per 10 cm plate.
- Daily or every 2 days reefed them. When they reach 80 to 85 % confluence, passage them. Or freeze.

4.4.1.2 Passage the ES cells

ES cells in healthy conditions generally reach 80-85% confluence in 2 to 4 days. Cells must be fed when the media turn to orange. When the media turn yellow, don't passage them. Before passaging the cells, feed them in the evening and the next morning before 3-4 hours.

- Aspirate the medium, wash one time with certain amount of PBS. Add certain amount of trypsin and incubate at 37^0C for certain time. Then same amount of ES cell media is added to inactivate trypsin.

	Amount of PBS	**Amount of trypsin**	**Incubation**
10 cm plate	10 ml	1ml	5 min
6-well-plate	3 ml	500μl	5 min
24-well-plate	1 ml	200 μl	3-4 min
48-well-plate	500 μl	100 μl	2-3 min
96-well-plate	100 μl	40 μl	2 min

- Again with pipetting, make them single cell. This is very critical step. Determine the cell numbers. Split them 1:3,1:4 or 1:10 and plate them into the suitable dishes.

4.4.1.3 Freezing the ES cells

Check the cells under the microscope for 80-85 % confluence. Refeed 3-4 hrs before passaging. Trypsinize cells from the 10 cm dish and add ES cell media to

inactivate as described. Put on ice, add same volume (1:1), 2 ml freezing media, then split into 1 ml freezing vials. Each vial should contain $3x10^6$ cell. For short-term storage put at -80 (overnight) or -150 (1-2 months). For long term storage, liquid nitrogen tanks are preferable.

4.4.2 Electroporation into ES cells

- Cell preparation is the critical step. 80% confluent ES cells should be split 1:2 one day before the electroporation. During the electroporation day, 3-4 hrs prior they should be reefeded.
- Trypsinize the cells and resuspend in media as described. Count 10 μl sample.
- Pellet the cells, and resuspend in PBS. ($1x10^7$ cells/900 μl PBS)
- 40μg linearized DNA(1μg/1μl TBE or PBS or water) is put onto the cell suspension, then transfer them to the gene pulser cuvvette (0.4 cm electrode gap)
- Incubate 5 min on ice
- Set electroporator 240 V, 500μF, and pulse the cells (time constant should read between 5.6-7.0).
- Leave the cuvvette on ice for 5 min. Then plate the $1x10^6$ cells in a 10 cm dish.
- After 24 hrs, with a suitable antibiotic, start colony selection. In this project, the cells after 24 hours were reefed by G418 (0.2μg/ml) and BSD (0.5μg/ml) for 3 days, then every 2 days the media was changed until the colonies were visible. For G418 around 10.day the colonies were ready to collect, for BSD after 14. day, the colonies were picked.

4.4.2.1 Picking the ES cells and transferring them to the 96 wells

The best time for picking the cell colonies is around 10 to 11 days, although they could be picked as late as 8 days, and colonies can be recovered 18 to 21 days after electroporation.

● When the time is suitable for picking the colonies, first remove the media and put around 4to 5 ml PBS on the 10 cm dishes. Under the microscope, using a 20μl micropipette, pick colonies into the 96 well plate.

● Through several pipetting up and down, make the colonies single cells. Then add 100μl ES cell medium containing suitable a selective reagent, antibiotic.
CRITICAL: The cells should be examined daily for their confluence. Generally in 2 days, the colonies reached certain confluence in 96 well plates and should be transfered into the 48 well plates. Another 2 days is required before transferring them to 24 well plates, and then in 2 days later transfer to 6 well plates through trypsinization and inactivation of the trypsin as described.

4.4.3 Transient transfection of DNA via lipid vesicles (Lipofectamine)

In this PhD project, this method was used to introduce FLEx-IRES-RFP constructs into mouse E14TG2a cells to quickly test the constructs and to transfect the pCAGGs-Cre-puro plasmid into the mutagenic oriented FLEx-IRES-RFP (with different version of pausing signals) random stables (mouse E14TG2acells). This protocol used in this project grew the cells in 6 well plates.

● Each 6 well plate contains $3x10^5$ cells/3ml medium. Mix 1μg DNA and 50μl DMEM (contains no serum) gently. Prepare lipofectamine, first mix the stock gently, and then put 3μl lipofectamine from the stock into the 50 μl DMEM (DNA:lipofectamine: 1:3) Incubate them at rt for 5 min.

● Following the incubation, mix 50 μl DMEM containing DNA and 50 μl DMEM containing lipofectamine. Incubate them at rt for 20 min.

● Then add 100 μl DNA and lipofectamine mixture to the 6 well, on top of the cells. Mix gently by rocking the plate back and forth.

● Incubate the cells at 37 5% CO_2 for 24 to 48 hours until they are ready to assay transgene expression. It is not necessary to change the medium.

5. MATERIALS

5.1 Chemicals and solutions

Acetic acid, glacial	**MERCK**
AA(Acrylamide/bisacrylamide)	**SEVERN BIOTECH LTD**
Agarose	**INVITROGEN,ROTH**
Ampicilline	**SIGMA**
APS	**SIGMA**
Bacto-Agar	**BD**
Blasticidin S HCl	**INVITROGEN (prokaryotic system)**
Blasticidin S	**INVIVOGEN (mammaliansystem)**
Bromophenolblue	**MERCK**
BSA	**SIGMA**
Chloramphenicol	**SIGMA**
Chloroform	**MERCK**
DMSO	**SIGMA**
ECL solution	**AMERSHAM**
EDTA disodium salt	**SIGMA**
Ethidiumbromide	**SIGMA**
Ethanol	**MERCK**
G418	**INVITROGEN**
Glycerol	**MERCK**
Glycogen	**ROCHE**
HCl	**MERCK**
IPGT	**SIGMA**
Isopropanol	**MERCK**
Kanamycin sulfate	**SIGMA**
KCl	**MERCK**
L-Arabinose	**SIGMA**
β-Mercaptoethanol	**SIGMA**
$MgCl_2$	**SIGMA**
Milkpowder	**HAIRLER**
Na_2HPO_4	**MERCK**
Na-citrate	**MERCK**
NaCl	**MERCK**
NaOAc	**MERCK**
NaOH	**MERCK**
Phenol/Chloroform/Isoamylalcohol	**SIGMA**
Puromycine	**SIGMA**
SDS	**BIORAD**
TEMED	**SIGMA**
Tetracycline	**SIGMA**

Trizma base	**MERCK**
Trizol	**SIGMA**
Tween 20	**SIGMA**
X-gal	**BIOMOL**

5.2 Media and Supplements

DMEM with Glutamax	**INVITROGEN**
FCS	**INVITROGEN**
L-Glutamine	**INVITROGEN**
LIF recombinant	**STEWART LAB**
Non essential aminoacids	**BIOCHROME**
Penicillin/Streptomycine	**INVITROGEN**
Sodium Pyruvate	**INVITROGEN**
Trypsin/EDTA 1x	**INVITROGEN**

5.3 Other reagents

5.3.1 Enzymes, markers and nucleotides and other reagents

All restriction enzymes	**NEB**
Asp718	**ROCHE**
Deoxyribonuclease I	**INVITROGEN**
Ribonuclease A	**SIGMA**
Ribonuclease H	**INVITROGEN**
Proteinase K	**MERCK**
Proteinase inhibitor coctail	**SIGMA**
T4 DNA ligase	**NEB**
Klenow large fragment	**NEB**
dNTP mix	**SIGMA**
1 kb DNA ladder	**INVITROGEN,NEB**
1 kb plus DNA ladder	**INVITROGEN**
Prestained protein marker,broad range	**NEB**
alpha ^{32}**P** dCTP	**AMERSHAM**

5.3.2 Kits

Plasmid maxi preparation kit	**QIAGEN**
Qiaquick PCR purification kit	**QIAGEN**
Qiaquick gel extraction kit	**QIAGEN**
BACMAX DNA purification kit	**EPICENTRE**
Triplemaster PCR system	**EPPENDORF**
High prime (random primed DNA labelling kit)	**ROCHE**
Lipofectamine 2000	**INVITROGEN**

5.3.3 Antibodies

Anti-beta actin monoclonal antibody	**SIGMA**
Mouse anti-GFP monoclonal antibody (unconjugated)	**ROCHE**
Anti-mouse Ig G conjugated horsedish preoxidase (goat)	**ROCHE**

5.4 Instruments

Balancer (TE3102S)	**SARTORIUS**
Cell culture centrifuge (Multifuge3 S-R)	**HERAEUS**
Cell counter CASY	**SCHARFE SYSTEMS**
Cell culture hood	**HOLTEN**
Cell culture incubator	**HERAEUS**
Cell culture incubator (Heracell 150)	**HERAEUS**
Cell culture microscope (CK40)	**OLYMPUS**
Cell culture fluorescent microscope(CKX41)	**OLYMPUS**
Cell culture rotor(6445)	**HERAEUS**
Cell culture water bath	**GFL**
Centrifuge table top(5415D)	**EPPENDORF**
Centrifuges	**BECKMAN COULTER(Avanti J25)**
Rotors JLA16.250,JS13.1	**BECKMAN COULTER**
Digital camera (C304OZoom)	**OLYMPUS**
ECL hyperfilm	**AMERSHAM**
Electrophoresis power supply	**EMBL PS143**
Electroporator bacterial cells(2510)	**EPPENDORF**
Electroporation cuvvette (1 mm)	**EPPENDORF**
Electroporator ES cells (gene pulser X)	**BIORAD**
Electroporation cuvvette (4 mm)	**BIORAD**
Geiger counter (LB122)	**BERTHOLD**
Gel documentation system	**BIORAD**
Glassware	**SCHUTT DURAN**

Hybridization oven	**BACHOFER**
Incubators (bacterial system)	**INFORS AG**
LAS3000 Image system	**FUJIFILM**
Nanodrop ND1000(spectrophotometer)	**PeqLab Biotechnologie GMBH**
PCR machine (Robocycler)	**STRATAGENE**
PCR machine (Epgradient S)	**EPPENDORF**
Semi-dry transfer blotter	**BIORAD**
Scintillation counter	**BECKMAN COULTER**
Spectrophotometer 2100	**AMERSHAM**
Spectrophotometer quartz cuvvette	**SIGMA**
Thermoshaker (1.5/2 ml)	**EPPENDORF**
UV stratalinker 2100	**STRATAGENE**
Vacuum oven	**KENDRO**

5.5 Disposables

Bacteria plates	**GREINER**
Cell culture freezing vials	**NUNC**
Cell culture pipettes (2,5,10,25 ml)	**COSTAR**
Cell scraper (23 cm)	**NUNC**
Centrifuge tubes (1.5 and 2 ml)	**EPPENDORFS**
Centrifuge tubes (12, 15 and 50 ml)	**FALCON**
Combitips	**EPPENDORFS**
Filters (0.22μm)	**MILLIPORE**
Filters (0.45μm)	**MILLIPORE**
Filters steritop (0.45μm, 500 ml)	**MILLIPORE**
Gloves	**SAENGER**
Inoculation needles	**NUNC**
Nitrocellulose membrane for protein transfer (protran)	**SCHLEICHER+SCHULER**
Positively charged nylone membrane for DNA blotting	**AMERSHAM**
PCR tubes (0.2 ml chain)	**SARSTEDT**
Pipette tips(P2,P10,P100,P1000)	**SARSTEDT and NERBE PLUS**
Spectrophotometer plastic cuvvettes	**SARSTEDT**
Syringes	**BECTON/DICKINSON**
Tissue culture plates (96,24,6 well)	**NUNC**
Tissue culture plates (48 well, 10 cm)	**FALCON**

5.6 Buffer Systems

5.6.1 Plasmid DNA purification system

P1 buffer	**P2 buffer**	**P3 buffer**
50 mM Tris-HCl	200 mM NaOH	3 M potassium acetate
10 mM EDTA	1% SDS	
100 μg/ml RNAse A		

QBT buffer	**QC buffer**	**QF buffer**
750 mM NaCl	1 M NaCl	1.25 M NaCl
50 mM MOPS	50 mM MOPS	50 mM Tris, pH 8.5
15% isopropanol	15% isopropanol	15% isopropanol
0.15% tritonX-100		

TE buffer

10 mM Tris

1mM EDTA

5.6.2 Agarose gel electrophoresis system

10X loading buffer	**TBE (5X)**
25% Ficoll	54 g Trizma base
100 mM EDTA	27.5 g Boric acid
Bromophenol blue	20 ml 0.5M EDTA, pH 8 add sterile water to 1 l

5.6.3 Genomic DNA purification system

Lysis buffer

50 mM Tris-HCl, pH 8

100 mM EDTA

100 mM NaCl

1% SDS

0.5 mg/ml Proteinase K(dissolve 100 mg proteinase K in 10 ml sterile water, aliquot and store at -20^{0}C).

5.6.4 Southern blotting system

Depurination solution		**Denaturation solution**		**Neutralization solution**	
37% HCl	10 ml	NaCl	87 g	NaCl	87 g
sterile water to	1l	NaOH pellets	20 g	Tris	121 g
		sterile water to	1 l	37% HCl	50 ml
				sterile water	to 1 l

20 X SSC

NaCl	175 g
tri-Sodium citrate dihydrate	88 g
sterile water to	1l

1M Hybridization stock		**1X Hybridization buffer**	
Na_2HPO_4	89 g	250 mM Na_2HPO_4	50 ml 1M stock
Adjust to pH 7.2 with H_3PO4		7% SDS	35 ml 20% stock
		1% BSA	1 g
		1mM EDTA	100 µl 0.5 M stock
		15 ml sterile water	

Wash buffer

20mM Na_2HPO_4, pH 7.2	40 ml	1M stock
1% SDS	50 ml	20% stock
1mM EDTA	2 ml	0.5 M stock
sterile water to	1 l	

20% SDS

200 g SDS+ 1 l water dissolve (heat)

5.6.5 Western blotting system

Protein extraction buffer (coctail)

250 mM Tris, pH 7.8	1 ml
1% protease inhibitors	10 μl
100 mM PMSF	10 μl
100 mM DTT	10 μl

Sample loading buffer

0.5 M Tris/HCl, pH 6.8

2% (v/v) β-mercaptoethanol

20% (v/v) glycerol

bromophenol blue in distilled water

Transfer buffer		Blocking buffer
39 mM glycine	2.9 g	5% milk powder
48 mM Trizma base	5.8 g	0.1% Tween 20 in PBS
0.037% SDS	1850 μl 20% SDS	
20% methanol	200 ml	
water to	1 l	

5.7 Cell Culture Systems

5.7.1 Bacterial culture system

For bacteria, small quantities of are usually grown on solid support that contains nutrients embedded on it called bacterial agar, and and large scale quantities of are grown in liquid media containing nutrients.

LB (Luria Bertani medium, pH 7.4)		**LB agar**	
1% bacto tryptone	10 g	15g	bacto agar
0.5 % bacto yeast	5 g	1l	LB,pH 7.4
1% NaCl	10 g		boil
water to	1 l		

Media for antibiotic selection

Chloramphenicol	**Stock (**30mg/ml in ethanol)	15 μg/ml for BAC and low copy plasmids, 50 μg/ml for high copy plasmids
Ampicillin	**Stock** (100mg/ml in 50% ethanol)	50 μg/ml for BAC and low copy plasmids, 100 μg/ml for high copy plasmids
Kanamycin	**Stock** (30mg/ml in sterile distilled water)	15 μg/ml for BAC and low copy plasmids, 50 μg/ml for high copy plasmids
Blasticidin S HCl	**Stock** (10mg/ml in sterile distilled water)	4 μg/ml for BAC and low copy plasmids, 5 μg/ml for high copy plasmids
Tetracycline	**Stock** (10mg/ml in 75% ethanol)	3 μg/ml for BAC and low copy plasmids, 5 μg/ml for high copy plasmids

5.7.2 Mammalian cell culture system

Basic media for mouse embryonic stem cells

Fetal calf serum (filtered through 0.22μm pore filter)	**75 ml**
100 U: μl/ml penicillin/streptomycin	**5 ml**
L-glutamine	**5 ml**
β-Mercaptoethanol (7μl in 10 ml PBS)	**5 μl**
Na pyruvate	**5 ml**
Non-essential aminoacids	**5 ml**
LIF (leukaemia inhibitory factor) 10^6 or 10^7 U	**1 ml**
DMEM glutamax	**400 ml**

2x Freezing media

50% Fetal calf serum

20% DMSO

in DMEM glutamax

Media for antibiotic selection

G418	**Stock** (50mg/ml in HEPES)	0.20 mg/ml
Blasticidin	**Stock** (5mg/ml in dd water)	5μg/ml
Puromycin	**Stock** (1mg/ml in dd water)	1μg/ml

6. REFERENCES

1. Abremski, K., Hoess, R. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein.
J. Biol. Chem. **259,** 1509-1514 (1984).

2. Ashe, H.L. et al. Intergenic transcription and transinduction of the human β-globin locus. Genes Dev. **11**, 2494-2509 (1997).

3. Ashfield, R. et al. MAZ dependent termination between closely spaced human complement genes. EMBO Journal **13**, 5656-5667 (1994).

4. Angelis, M.H., Adler, A. et al., Imai, K. et al Mouse Genomics.
In The laboratory mouse: the handbook of experimental animals (e.d Hans Hedrich) pp 47-83. Elsevier Academic Press USA (2004).

5. Austin, S., Ziese, M., Sternberg, N. A novel role for site-specific recombination in maintenance of bacterial replicons.
Cell **25,** 729-736 (1981).

6. Bahramian, M.B. Coupling of transcription termination to RNAi.
Journal of Theoretical Biology **245**, 278-289 (2007).

7. Bentley, D.L. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Current Opinion in Cell Biology **17**, 251-256 (2005).

8. Berger, W. et al. An animal model for Norrie disease (ND): gene targeting of the mouse ND gene. Human Molecular Genetics **5,** 51-59 (1996).

9. Berget, S.M. Exon recognition in vertebrate splicing
The Journal of Biological Chemistry **270**, 2411-2414 (1995).

10. Bollag,R.J. et all. Homologous recombination in mammalian cells.
Annu.Rev.Genet. **(23),** 199-225 (1989).

11. Branda, K.S., Dymecki, S.M. Talking about revolution: The impact of site-specific recombinases on genetic analyses in mice. Developmental Cell **6,** 7-28 (2004).

12. Brault, V. et al. Modeling chromosomes in mouse to explore the function of genes, genomic disorders, and chromosomal organisation.
PLoS Genetics **2**, 911-919. (2006).

13. Bucholz, F., Ringrose, L., Angrand, P.O., Rossi, F., Stewart, A. F. Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. Nucleic Acids Res. **1,** 4256-4262 (1996).

14. Bucholz, F., Angrand, P.O., Stewart, A.F. Improved properties of FLP recombinase evolved by cycling mutagenesis. Nat. Biotechnol. **16,** 657-662 (1998).

15. Buratowski,S. Connections between mRNA 3^{I} end processing and transcription termination. Current Opinion in Cell Biology **17**,257-261 (2005).

16. Capecchi, M.R. Generating mice with targeted mutations.
Nature Medicine **7,** 1086-1090 (2001).

17. Carlson, C.M., Largaespada, D.A. Insertional Mutagenesis in mice: new perspectives and tools. Nature Reviews Genetics **6,** 568-580 (2005).

18. Chen F., MacDonald, C., Wilusz, J. Cleavage site determinants in the mammalian polyadenylation signal. Nucleic Acids Res. **23**, 2614-2620 (1995).

19. Colgan,D.F., Mantley, J.M. Mechanism and regulation of mRNA polyadenylation. Genes Dev. **11**, 2755-2766 (1997).

20. Copeland, N.G. et al. Recombineering: a powerful new tool for mouse functional genomics. Nature Reviews Genetics **2,** 769-779 (2001).

21. Craig, N.L. The mechanism of conservative site-specific recombination.
Annual Rev. Genet. **22,** 77-105 (1988).

22. Dahme, M. et al. Distruption of the mouse L1 gene leads to malformations of the nervous system. Nature Genet.**17**, 346-349(1997).

23. Deng, C., Capecchi, M.R. Re-examination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. Mol.Cell Biol. **12,** 3365-3371 (1992).

24. Dominguez-Bendala, J. Elevated expression of exogenous Rad51 leads to identical increases in gene-targeting frequency in murine embryonic stem (ES) cells with both functional and dysfunctional p53 genes. Experimental Cell Research **(286),** 298-307 (2003).

25. Dye, M.J., Proudfoot, N.J. Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. Molecular Cell **3**, 371-378 (1999).

26. Dye,M.J., Proudfoot, N.J. Multiple transcriptional cleavage precedes polymerase release in termination by RNA Polymerase II. Cell **105**, 669-681 (2001).

27. Dymecki, S.M. In Gene targeting: a practical approach
(e.d A.L Joyner) pp.37-99. IRL press Oxford (2000).

28. Eggleston, A.K., West, S.C. Exchanging partners: Recombination in E.coli.
Trends Genet. **12,** 20-26 (1996).

29. Ellerstrom, C. et all. Facilitated expansion of human embryonic stem cells by single cell enzymatic dissociation. Stem Cells,1-37(2007).

30. Evans, M.J. Gene trapping- a preface. Developmental Dynamics **212,** 167-169 (1998).

31. Favard, S. et all. Electrotransfer as a non-viral method of gene delivery Current Gene Therapy **(7)**, 67-77 (2007).

32. Fox-Walsh, K.L. et al. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc. Natl. Acad. Sci. **102**, 16176-16181 (2005).

33. Friedel, R.H. et al. Gene targeting using a promoterless gene trap vector ("targeted trapping") is an efficient method to mutate a large fraction of genes. Proc. Natl. Acad. Sci. **102**, 13188-13193 (2005).

34. Friedrich, G., Soranio, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev. **5,** 1513-1523 (1991).

35. Gossler, A.J., Joyner, A.L, Rossant, J., Skarners, W.C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science **244**, 463-465 (1989).

36. Gromak et al. Pause sites promote transcriptional termination of mammalian RNA Polymerase II. Mol.Cell Bio. **26**,3986-3996 (2006).

37. Guenet, JL., Bonhomme, F. Origin of the laboratory mouse and related subspecies. In The laboratory mouse: the handbook of experimental animals (e.d Hans Hedrich) pp 3-13. Elsevier Academic Press USA (2004).

38. Hamilton, D.L., Abremski, K. Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre mediated synapsis of two lox sites.
J. Mol. Bio **178**, 481-486 (1984).

39. Hansen, J. et all. A large scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. Proc. Natl. Acad. Sci. **(117),** 9918-9922 (2003).
40. Hasty, P., Ramirez-Solis, R., Krumlauf, R., Bradley, A. Introduction of a subtle mutation into the Hox 2.6 locus in embryonic stem cells. Nature, **350**,1-4 (1991).

41. Hasty, P., Bradley, A. In Gene targeting: a practical approach
(e.d A.L Joyner) pp.1-35. IRL press Oxford (2000).

42. Heber, S. et al. Mice with combined gene knock-outs reveal essential and partially redundant functions of amyloid precursor protein members.
The Journal of Neuroscience **20,** 7951-7963 (2000).

43. Hoess, R.H., Abremski, K. Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. Proc. Natl. Acad. Sci. **81,** 1026-1029 (1984).

44. Hoess, R.H., Abremski, K. Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. J. Mol. Bio **181,** 351-362 (1985).

45. Houdabine, M.L.The mouse as an animal model for human diseases.
In The laboratory mouse: the handbook of experimental animals (e.d Hans Hedrich) pp 97-110. Elsevier Academic Press USA (2004).

46. Hrabe de Angelis, M.H. Genome-wide, large scale production of mutant mice by ENU mutagenesis. Nature Genetics **25**, 444-447 (2000).

47. International Gene Trap Consortium. A public gene trap resource for mouse functional genomics. Nature Genetics **36**, 543-544 (2004).

48. Kim, D.G et all. The construction of a bifunctional mRNA in the mouse by using internal ribosomal entry site of the encephalomyocarditis virus.
Mol. Cell Bio. **12**, 3636-3643 (1992).

49. Koonin, E.V.et al. The impact of comparative genomics on our understanding of evolution. Cell **101,** 573-576 (2000).

50. Korn, R. et al. Enhancer trap integrations in mouse embryonic stem cells give rise to staining patterns in chimaeric embryos with a high frequency and detect endogeneous genes. Mech.Dev.**39,** 95-109 (1992).

51. Lakso, M. et al. Targeted oncogene activation by site-specific recombination in transgenic mice. Proc. Natl. Acad. **89,** 6232-6236 (1992).

52. Langui D., Lachapelle, F., Duyckaerts, C. Animal models of neurodegenerative diseases. Med. Sci. (Paris) **23**, 180-186 (2007).

53. Lee, G., Saito, I. Role of nucleotide sequences of lox P spacer region in Cre-mediated recombination. Gene **216**, 55-65 (1998).

54. Legendre, M., Gautheret, D. Sequence determinants in human polyadenylation site detection.BMC Genomics **4**, 1-9 (2003).

55. Levitt, N., Brigss, D., Gil, A., Proudfoot , N.J. Definition of an efficient synthetic poly(A) site. Genes Dev. **3**, 1019-1025 (1989).

56. Logan, J. et al. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse β^{maj} -globin gene. Proc.Nat.Acad.Sci. **84**, 8306-8310 (1987).

57. Lou, H., Gagel, R.F., Berget, SM. An intron enhancer recognized by splicing factors activates polyadenylation. Genes Dev. **15**, 208-219 (1996).
58. Lou, H. et al. Regulation of alternative polyadenylation by U1 snRNPs and SRp20. Mol. Cell Biol. **18**, 4977-4985 (1998).

59. MacDonald, C., Wilusz, J., Shenk,T. The 64- kilodalton subunit of the CstF polyadenylation factor binds to premRNAs downstream of the cleavage site and influences cleavage site location.Mol. Cell Biol. **14**, 6647-6654 (1994).

60. Manivasakam, P. et all. Restriction enzymes increase efficiencies of illegitimate DNA integration but decrease homologous integration in mammalian cells. Nucleic Acids Res. **(29),** 4826-4833 (2001).

61. Matsuzaki, H. et al. Gene conversation associated with site-specific recombination in yeast. Mol.Cell Biol. **8,** 955-962 (1988).

62. Matzuk, M.M., et al. Alpha-inhibin is a tumor suppressor gene with gonadal specificty in mice. Nature **360,** 313-319 (1992).

63. Mouse Genome Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature **420,** (2002).

64. Muller, U. Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. Mech.Dev.**82,** 3-21(1999).

65.Mulley, J., Holland, P. Comparative genomics: small genome, big insights.
Nature 431, 916-917 (2004).

66. Murphy, K.C. Use of bacteriophage λ recombination functions to promote gene replacement in Escherichia coli. J.Bacteriol. **180**, 2063-2071 (1998).

67. Muyrers, J.P., Zhang, Y., Testa, G., Stewart, A.F. Rapid modification of bacterial artificial chromosome by ET cloning. Nucleic Acids Res. **27,** 1554-1557 (1999).
68. Muyrers, J.P.P., Zhang, Y., Stewart, A.F. ET cloning: Think recombination first. In Genetic Engineering, Principles and Methods (ed. J.K. Setlow)**.** pp. 77-98 Kluwer Academic/Plenum Publishers, USA,2000.

69. Muyrers, J.P.et all. RecE/RecT and Redα/ Redβ initiates double-stranded break repair by specifically interacting with their respect partners.
Genes and Development **14**, 1971-1982 (2000).

70. Muyrers, J.P. et all. Point mutation of bacterial artificial chromosomes by ET recombination. EMBO Reports **1**, 239-243 (2000).

71. Nagy, A. et al. Tailoring the genome: the power of genetic approaches.
Nature Genetics Supplement **33**, 276-284 (2003).

72.National Human Genome Research Institute Review Series (www.ornl.gov).

73. Neugebauer, K.M. On the importance of being co-transcriptional. Journal of Cell Science **115**, 3865-3871 (2002).

74. Nishikura, K. Editor meets silencer: crosstalk between RNA editing and RNA interference. Nat.Rev.Mol.Cell Bio.**7**, 919-931 (2006).

75. Niwa, M., Macdonald, C.C., Berget, S.M. Are vertebrate exons scanned during splice site selection? Nature **360**, 277-280 (1992).

76. Nolan, M.P et al. A systematic, genome-wide, phenotype-driven mutagenesis. Nature Genetics 25, 440-443 (2000).

77. Nord, A.S. et al. The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse.
Nucleic Acid.Res.**34,** 642-648 (2006).
78. Olson, E.N et al. Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene Mrf4. Cell **85**, 1-4 (1996).

79. Orban, P.C., Chui, D., Marth, J.D. Tissue and site-specific DNA recombination in transgenic mice. Proc. Natl. Acad.Sci. **89**, 6861-6865 (1992).

80. Peterson, M.L., Bertolino, S., Davis, S. An RNA polymerase pause site is associated with the immunoglobulin μs poly(A) site. Mol.Cell Biol. **22,** 5606-5615 (2002).

81. Plant, K.E. et al. Strong polyadenylation of weak pausing combine to cause efficient termination of transcription in human $^{G}\gamma$-globin gene.
Mol. Cell Bio. **25**, 3276-3285 (2005).

82. Proudfoot, N.J. Genetic dangers in poly(A) signals. EMBO Reports **2**
891-892 (2001).

83. Proudfoot, N.J. et al. Integrating mRNA processing with transcription.
Cell, **108**, 501-512 (2002).

84. Rautmann, G. et al Synthetic donor and splice acceptor splice sites function in an RNA polymerase B(II) transcription unit. EMBO **3,** 2021-2028 (1984).

85. Raymond, C.S., Soranio, P. Engineering mutations: deconstructing the mouse gene by gene. Developmental Dynamics **235,** 2424-2436 (2006).

86. Ringrose, L. et al. Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances.
EMBO Journal, **18**, 6630-6641 (1999).

87. Robberson, B.L., Cote, G.J., Berget, S.M. Exon definition may facilitate splicesiteselection in RNAs with multiple exons. Mol.Cell Biol. **10,** 84-94 (1990).
88.Rossant, J., Scherer, S.W. The mouse genome sequence- the end of the tail, or just the beginning. Genome Biology **4,** 109.1-109.4 (2003).

89. Sauer, B., Henderson, M. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc. Natl. Acad. Sci. **85,** 5166-5170 (1988).

90. Schimenti, J., Bucan,M. Functional genomics in the mouse: phenotype-based mutagenesis screens. Genome Research **8,** 698-710 (1998).

91. Schnutgen,F. et all. Genomewide production of multipurpose alleles for functional analysis of mouse genome. Proc. Natl. Acad. Sci. **(102),** 7221-7226 (2005).

92. Sipos, L., Gyurkovics, H. Long-distance interactions between enhancers and promoters.The case of the Abd-B domain of the Drosophila bithorax complex.
FEBS Journal **272,** 3253-3259 (2005).

93. Smithies, O. et al. Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. Nature **317**, 230-234 (1985).

94. Skarnes, W.C. Two ways to trap a gene in mice. Proc. Natl. Acad. Sci. **102**,13001-13002 (2007).

95. Soewarto, D., Random ENU mutagenesis. In Transgenic mouse: methods in molecular biology.**209** pp249-266. Humana Press Inc. USA 2003.

96. Sreenath, T. et al. Dentin sialophosphoprotein knock out mouse teeth display widened predentin zone and developdefective dentin mineralization similar to human dentinogenesis imperfecta type III. JBC **278,** 24874-24880 (2003).

97. Stanford, W.L., Cohn, J.B., Cordes, S.P. Gene trap mutagenesis: past, present and beyond. Nature Reviews Genetics **2,** 756-768 (2001).

98. Sung, Y.H. et al. Functional genomics approach using mice. J.Biochem.Mol.Biol. **37**, 122-137 (2004).

99. Tanaka, K. et al. Epilepsy and exacerbation of braini njury in mice lacking the glutamate transporter GLT-1. Science **276**, 1699-1702 (1997).

100. Testa, G. et al. BAC engineering for the generation of ES cell targeting construct and mouse transgenes. Methods Mol.Bio. **256**, 123-139 (2004).

101. The International Mouse Knockout Consortium. A mouse for all reasons. Cell**128**, 9-13. (2007).

102. Thomas, JW. et al. X ray induced mutations in mouse embryonic stem cells. Proc. Natl. Acad. Sci **95**, 1114-1119 (1998).

103. Tian, B. et al. A large scale analysis of mRNA polyadenylation of human and mouse genes
Nucleic Acids Res. **33,** 201-212 (2005).

104. To, C., Stanford,W.L. The centre for modelling human disease gene trap resource. Nucleic Acids Res. **32,** 557-559 (2004).

105. Wells, C., Brown, S.D. Genomics meets genetics.
Mamm. Genome **11,** 472-477 (2000).

106. West, A.G., Gaszner, M., Felsenfeld, G. Insulators: many functions, many mechanisms. Genes&Dev. **16,** 271-288 (2002).

107. West, S. et al. Transcription termination sequences in the mouse serum albumin gene. RNA **12**, 655-665 (2006).

108. Williams, R.W. et al. The math of making mutant mice.
Genes, Brain and Behaviour **2**, 191-200 (2003).

109. Wu, H., Liu, X., Jaenisch, R. Double replacement: Strategy for efficient introduction of subtle mutations into the murine Col1a-1 gene by homologous recombination in embryonic stem cells.
Proc. Natl. Acad. Sci. **91,** 2819-2823 (1994).

110. Wurtele,H. et all. Illegitimate DNA integration in mammalian cells.
Gene Therapy **(10),** 1791-1799 (2003).

111. Wurst, W., Gossler, A. Gene trap strategies in ES cells. In Gene targeting: a practical approach
(e.d A.L Joyner) pp.207-254. IRL press Oxford (2000)

112. Vasquez, K.M. et all. Manipulating the mammalian genome by homologous recombination
Proc. Natl. Acad. Sci. **(98),** 8403-8410 (2001).

113. Vishvanath, R.I. Translating genome sequences into biological understanding. Genome Biology **4:324** Meeting report (2003).

114. von Melchner, H. et al. Selective distruption of genes expressed in totipotent embryonal stem cells. Genes Dev. **6,** 919-927 (1992).

115. Voss, A.K., Thomas,T., Gruss,P. Efficiency assessment of the gene trap approach. Developmental Dynamics **212,** 171-180 (1998).

116. Yonaha, M., Proudfoot, N.J. Transcriptional termination and coupled polyadenylation in vitro. EMBO Journal **19**, 3770-3777 (2000).

117.Yonaha, M., Proudfoot, N.J. Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. Molecular Cell **3**, 593-600 (1999).

118. You, Y. Chromosomal deletion complexes in mice by radiation of embryonic stem cells. Nature Genetics **15,** 285-288(1997).

119. Zambrowicz, B.P., Friedrich, G.A., Comprehensive mammalian genetics: history and future prospects of gene trapping in the mouse. Int.J.Dev.Biol.**42,** 1025-1036 (1998).

120. Zaratiegui, M. et al. Noncoding RNAs and gene silencing.
Cell **128**, 763-776 (2007).

121. Zaret, K.S., Sherman, F. DNA sequence required for efficient transcription termination in yeast. Cell **28**,563-573 (1982).

122. Zhang, Y., Bucholz, F., Muyrers, J.P., Stewart, A.F. A new logic for DNA engineering using recombination in E.coli. Nature Genetics **20,** 123-128 (1998).

123. Zhang, Y., Muyrers, J.P., Stewart, A.F. DNA cloning by homologous recombination in E.coli. Nature Biotech. **18,** 1314-1317 (2000).

124. Zheng, B, Mills, A.A, Bradley,A. A system for rapid generation of coat color-tagged knock-outs and defined chromosomal rearrangements in mice.
Nucleic Acids Res. **27**, 2354-2360 (1999).

125. Zheng, B. et al. Engineering mouse chromosomes with Cre-loxP: Range, efficiency, and somatic applications.Mol.Cell.Bio, **20,** 648-655 (2000).

126. Zhu, X.D., Sadowski, P.D. Cleavage-dependent ligation by the FLP recombinase. Characterization of a mutant FLP proteinwith and alteration in a catalytic aminoacid. J. Biol.Chem. **270**, 23044-23054 (1995).

127. Zimmer, A. Manipulating the genome by homologous recombination in embryonic stem cells.
Annu.Rev.Neurosci. **(15),** 115-137 (1992).

128. Xin, H-B., et all. Gene trap and gene inversion methods for conditional gene inactivation in the mouse. Nucleic Acids Res. **(33),** 1-10 (2005).

7. ABBREVIATIONS

APLP	Amyloid precursor like protein
BAC	bacterial artificial chromosome
BSD	blasticidin
bp	base pairs
CMV	Cytomegalovirus
ColA1	Collagen alpha 1
CoTC	co-transcriptional cleavage
C terminus	carboxy terminus
DNA	deoxyribonucleicacid
EDTA	ethylene diamine tetra acetic acid
GLT1	Glutamate transporter
hprt	hypoxanthine-guanine phosphoribosyl transferase
kb	kilobase
kDA	kilodalton
MRFP	Myogenic regulatory factor
mRNA	messenger ribonucleic acid
Myf5	Myogenic factor 5
ND	Norrie disease
neo	neomycine
o/n	overnight
pA	polyadenylation
PCR	polymerase chain reaction
PGK	phosphoglycerate kinase
RNA	ribonucleic acid
RT	room temperature
sA	splicing acceptor
wt	wild-type

Acknowledgement

First of all, I am grateful to my thesis supervisors Prof. Francis Stewart and Dr. Konstantinos Anastassiadis for sharing their valuable experiences in this interesting scientific area, and providing insight, advice, support, critics and knowledge during my PhD study time.

I would like to thank my thesis review and exam committee including Prof. Francis Stewart, Prof. Dr. Med. Christian Thiede, Prof. Dr. Harald von Melchner, Prof.Dr.Bernard Hoflack, and Assoc.Prof. Dr. Frank Bucholz for their valuable directions and discussions.

My special thanks to go our University Rector,Herr Prof.Dr. Herman Kokenge, PhD advisors Herr Rimmele and Frau Dicke for their trust, patience and support.

I would like to special thanks also the brains at the back side of the search engine GoPubMed and MeshPubMed that provided my thesis reviews, and all scientific papers. Prof.Schroder, Dr. Michael R. Alvers, Andreas Domms and all Transinsight Gmbh members. Studying with you was a big joy.

And my beautiful family, my mother Sevtap, father Atila and my brother Orhan without your supports, love, I could not handle the difficult times. This thesis is your credit, your promotion. My dear Dodo, Zeynep I heard your wishes and I hanged on always, this was the fruit.

Short time before my DAAD scholarship acceptance, dear Falk Sobiele, dear Anne Chesneau, I never forget your precious support.

Bobritz family and dear Martin Bobritz. You are my family in Germany and this thesis was built with your wonderful support and blessings.

My dear second mummy Saniye, I will always remember you in a lovely way. Thank you.

My dear friends Ceren Baki and Evren Koban, I started my MSc thesis with you, now I finished my PhD thesis with you. I learned long-life friendship is everything.

When the time comes to thank my lab gang! First of all dear Cagri, I believed God exist and gave you patience about my hours and hours conversations. And our handsome Italian man Roberto, thank you to push me write immediately this thesis and share your nice laughing, jokes, conversations. And dear Maria, I loved every seconds of our lunch times, dance activities and discussing our experiment problems. Wish to meet you before. Dear Giovanni, short before handling this thesis you came and gained my heart, millions of thanks your encouragement. Dear Assen, yes I took everything serious as we discussed otherwise this job could not be done properly. Hey Misho, if you were not there, if our light fighting were not there, this though PhD research time would be boring. Dear Guiseppe, only 10 minutes discussion after my group meeting in your office brought me in this level, thank you. And Luke, I loved every minutes of our evening discussions before my group meetings. Thank you for your nice, unforgettable friendship in this research group. And I would like to thanks Susann Schneider and Jingwen Wolkersdorf to share their technical experiences. Axel, last minute you made me added your name, thank you for German translation of my thesis headline.

I gladly wish to extend my gratitude to Dr. Youming Zhang for his kindness as well as his optimisim and strong biology knowledge to translate into the practical.

And my secret advisor, my close friend Andrea Kranz, I knew the first day we could work together, discuss. First of all I am thankful for your sharing your hours, every minutes, even seconds for pre-revewing my thesis in addition to your experiments and other responsibilities. I am grateful also to be your friend. You, I and Maria will stay in scientific area as spice mice members for long years. You will be genuine group leader.

And all the Prof Schroder Bioinformatics group members, and my dear party fans Gihan, Joerg, Dimitra, Heiko, Andreas, Annalisa. If you did not support me, and make me remember life is beautiful during the party times, barbecues and every lunch times, this book could not be written. By the way as a promise to keep, this book also was dedicated to Gihan and Dimitra.

Dear Prof. Wieland Hutner, it is real pleasure and luck to meet you, and my dear friends in this group, Jifeng and Davide.

Dear Emral Cakmak all the difficult times, you always standed near me. Seni daima sakin, magrur ama sevgi dolu bir dostum olarak hatirlayacagim. Dear Yigit, promise I never wear these blue stilettos. Thank you to be with me for the final stage.

And the last thanks to go my Chevalier. Nothing is enough to explain my happiness about your power in my life. You believed my capability, and made me believe. Always know, writing was joyful because of you and your cheerful spirit. With my sun clock, I caught the thesis submission deadline. This thesis belonged to you too.

Bu tez gece ve gunduz demeden dunya barisi icin canini yola koymus KAHRAMAN MEHMETCIGE adanmistir. Varolun.